Investigating the functional significance of an FGFR2 intronic SNP in Breast Cancer

Robbez-Masson, Luisa

Investigating the functional significance of an *FGFR2* intronic SNP in Breast Cancer

Luisa Robbez-Masson

Submitted in partial fulfillment of the requirements for
the Degree of
Doctor of Philosophy

May 2013

Queen Mary University
Centre for Tumour Biology
Barts Cancer Institute
John Vane Science Centre, Charterhouse Square
EC1M 6BQ London, UK

Declaration of authorship

I hereby declare that the material presented in this thesis is the result of original work done by the author, Luisa Robbez-Masson, at the Centre for Tumour Biology, Barts Cancer Institute, Queen Mary University of London. All the external sources have been properly acknowledged.

Acknowledgments

I would like to express my extreme gratitude to my supervisor Dr Richard Grose and thank him for making me a member of his research group, giving me well needed guidance, advice and encouragement throughout my three years PhD. I would like to thank my dear friends Dr Athina-Myrto Chioni, Stacey Coleman and Abbie Fearon for being great lab mates and friends and making the FGF group such a special place to work in.

My gratitude also goes to Prof Ian Hart, for leading the Centre for Tumour Biology with such enthusiasm and providing us with high standards of work ethic and great motivation. Thank you also to Dr Jude Fitzgibbon and Dr Csaba Bodor for providing much advice on SNP genotyping and reagents. Thank you to Prof Helen Hurst for her advice and input on ER biology and molecular biology techniques. I owe many thanks to the people of Prof Bruce Ponder's lab in CRUK, for their help and advice on ChIP, particularly Dr Kerstin Meyer and Dr Martin O'Reilly. I would like to acknowledge all of my colleagues from Barts Cancer Institute and office G28, especially Hector, Müge, Wasfi, Mo, and Carine.

I will always be indebted to my partner Yoann, for his support and encouragement throughout my PhD, who has helped me maintain my focus, be more optimistic and believe in myself and without whom I would not have completed this thesis.

I also want to thank my family in France, especially my parents Cathie and Frank as well as my sister Victoria and brother Lucas for their support, care and love.

Lastly I would like to thank Queen Mary University for the funding provided which allowed me to pursue my dream of starting an academic career in London.

"Somewhere, something incredible is waiting to be known."
Carl Sagan

"We are survival machines – robot vehicles blindly programmed to preserve the selfish molecules known as genes. This is a truth which still fills me with astonishment."

The selfish gene
Richard Dawkins

"Behind every man now alive stand thirty ghosts, for that is the ratio by which the dead outnumber the living. Since the dawn of time, roughly a hundred billion human beings have walked the planet Earth.

Now this is an interesting number, for by a curious coincidence there are approximately a hundred billion stars in our local universe, the Milky Way. So for every man who has ever lived, in this Universe there shines a star.

But every one of those stars is a sun, often far more brilliant and glorious than the small, nearby star we call the Sun. And many--perhaps most--of those alien suns have planets circling them. So almost certainly there is enough land in the sky to give every member of the human species, back to the first ape-man, his own private,
world-sized heaven—or hell.

2001: A Space Odyssey
Arthur C. Clarke

Abstract

Single nucleotide polymorphisms present in the second intron of the *fibroblast growth factor receptor 2* (*FGFR2*) gene have been linked with increased risk of breast cancer in several genome wide association studies. The potential effect of those SNPs appeared to be mediated through the differential binding of *cis*-regulatory elements, such as transcription factors, since all the SNPs in linkage disequilibrium were located in a regulatory DNA region. Preliminary studies have shown that a Runx2 binding site is functional only in the minor, disease associated allele of rs2981578, resulting in increased expression of *FGFR2* in cancers from patients homozygous for that allele. Moreover, the increased risk conferred by the minor *FGFR2* allele is associated most strongly in oestrogen receptor alpha positive (ERα) breast tumours, suggesting a potential interaction between ERα and FGFR signalling. Here, we have developed a human cell line model system to study the effect of those SNPs on cell behaviour. In an ERα positive breast cancer cell line, rs2981578 was edited using Zinc Finger Nucleases. Unexpectedly, the acquisition of the single risk allele in MCF7 cells failed to affect proliferation or cell cycle progression. Binding of Runx2 to the risk allele was not observed. However FOXA1 binding, an important ERα partner, appeared decreased at the rs2981578 locus in the risk allele cells. Additionally, differences in allele specific expression (ASE) of *FGFR2* were not observed in a panel of 72 ERα positive breast cancer samples. Thus, the apparent increased risk of developing ERα positive breast cancer is not caused by rs2981578 alone. Rather, the observed increased risk of developing breast cancer might be the result of a coordinated effect of multiple SNPs forming a risk haplotype in the second intron of *FGFR2*.

Table of Contents

List of Figures, Tables and Appendices

List of Abbreviations

BSA: Bovine serum albumin

BRCA1 and 2: Breast cancer type 1 and 2 susceptibility proteins

cDNA: complementary DNA

Cbl: named after Casitas B-lineage Lymphoma

ChIP: Chromatin Immunoprecipitation

CNV: Copy number variation

DAG: Diacylglycerol

DAPI: 4', 6-diamidino-2-phenylindole

eQTL: Expression quantitative trait loci

ERα: Oestrogen Receptor alpha

ERK: Extracellular signal-regulated kinases

FBS: Foetal bovine serum

FGF: Fibroblast growth factor

FGFR: Fibroblast growth factor receptors

FRS2: Fibroblast growth factor receptor substrate 2

g: Gravitational force

GAPDH: Glyceraldehyde phosphate dehydrogenase

gDNA: genomic DNA

GRB2: Growth factor receptor-bound protein 2

GWAS: Genome wide association study

HPRT: Hypoxanthine-guanine phosphoribosyltransferase

HR: Homologous recombination

HSPG: Heparan sulphate proteoglycans

IRES: Internal ribosome entry site

JAK: Janus kinase

LD: Linkage disequilibrium

MAPK: Mitogen-activated protein kinases

NHEJ: Non-homologous end joining

nm: nanometer

Oct1: Octamer transcription factor (also called POU2F1)

PBS: Phosphate buffered saline

PFA: Paraformaldehyde

PI3K: Phosphatidylinositol 3-kinase

PiP3: Phosphatidylinositol (3,4,5)-triphosphate

PKC: Protein kinase C

PLCγ: Phosphoinositide phospholipase C

RT: Room temperature

RTK: Receptor tyrosine kinase

Runx2: Runt-related transcription factor 2

SDM: Site directed mutagenesis

Sef: Similar expression to FGF

SEM: Standard error of the mean

SNP: Single nucleotide polymorphism

SPRY2: Sprouty homolog 2

STAT: Signal Transducer and Activator of Transcription

ZFN: Zinc-finger nuclease

CHAPTER 1

INTRODUCTION

1. Introduction

1.1. Breast Cancer

1.1.1. Incidence and mortality

Breast cancer is the most common cancer in women in the UK with 48,700 new cases diagnosed in 2009 (Data and Statistics, 2012). It is now the second most common cause of cancer-related death in European women, after lung cancer (WHO, 2008). Between 1971 and 2010, the incidence rate has increased by 90%, while overall mortality has decreased, however different trends can be observed over this period (Fig. 1.1A).

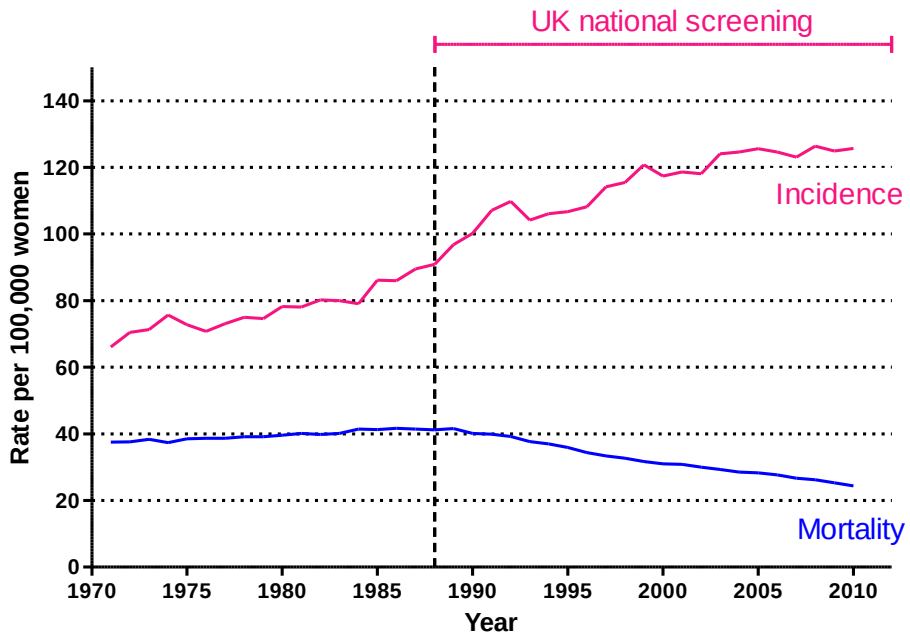
Female breast cancer age-standardised incidence rate increased steadily by around 1 to 2% each year from the mid-1970s (Data and Statistics, 2012). The apparent increase between 1980 and 1987 coincides with increased mammographic screening, with a national screening programme introduced in the UK in 1988 for women aged between 50 and 64 years old, following the recommendations of the Forrest Committee (Forrest P, 1986). By the mid-1990s, the increase in incidence rates returned to the pre-screening pace and continued this way until the mid 2000s, after which time the rate has remained relatively stable. The decrease in mortality observed from 1990 can be attributed both to improvement in treatments and the benefits of large scale screening, i.e. early detection.

Breast cancer can also affect males but is much less common, affecting just one in 100,000 men in England (Gomez-Raposo *et al*, 2010). This represents about one man for every 130 women diagnosed in the UK.

1.1.2. Pathology and progression

Like most solid cancers, primary breast cancers usually (but not necessarily) progress from local disease that can originate from the breast lobules or ducts (Lobular and Ductal carcinoma *in situ*, LCIS and DCIS respectively) to more invasive disease (Stages I to IV) leading to systemic metastasis (Fig. 1.1B) (Kufe *et al*, 2003). However, the natural history and prognosis vary considerably from patient

A



B

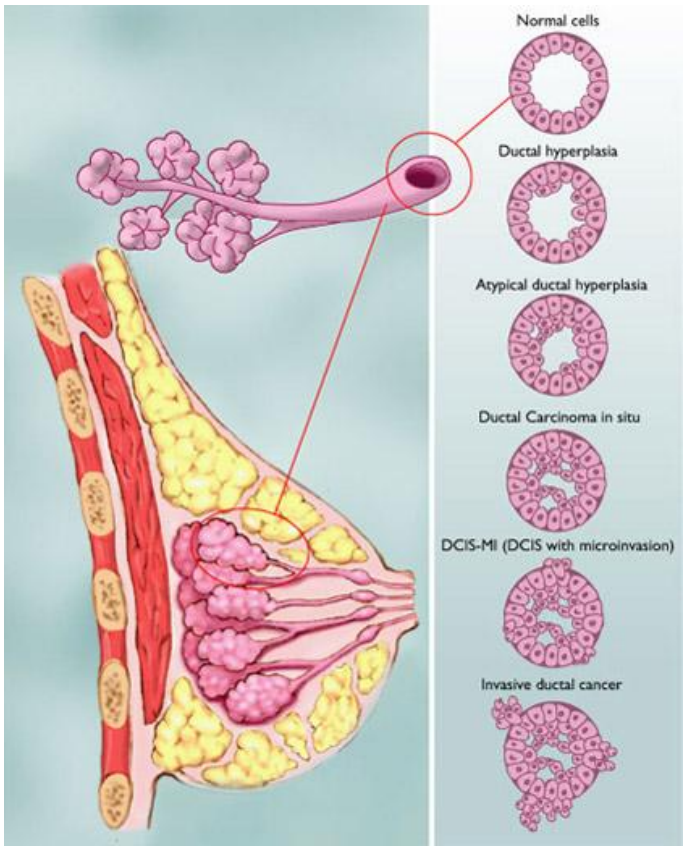


Figure 1.1: Breast cancer incidence and normal breast with ductal carcinoma *in situ* (DCIS) progression

A) Female breast cancer, European Age-Standardised Incidence and Mortality Rates, England from 1971 to 2010. UK National screening began in 1988 as represented by the vertical dashed line. B) General breast morphology and early stages of breast cancer development. Source: Data from UK National Statistics (2012) and image from breastcancer.org (Bryson, 2012).

to patient. Some patients have a very slow progressing disease that can be cured by local therapy and can survive for many years even after developing metastases. Historically, it even was established that a small percentage of patients survived more than 10 years without any treatment (Kufe *et al*, 2003). In other patients, the disease follows an aggressive, rapidly progressing course that is refractory to treatment (e.g. Triple-negative and Basal-like tumours) (Kufe *et al*, 2003).

Early-stage invasive breast cancer can be managed successfully with either mastectomy or breast conservation therapy (Fisher *et al*, 1989; Fisher *et al*, 2002). The results of the long term (20 years) follow up study by Fisher *et al* (2002) showed that additional adjuvant chemotherapy (hormonal or cytotoxic) or radiotherapy in surgery patients, implemented in the 1980s, led to significantly lower recurrence rates compared to those who underwent surgery alone. In addition, axillary lymph node staging (presence of metastasis in the axillary lymph nodes, particularly in the sentinel lymph node) is a powerful prognostic factor (Krag *et al*, 2010), and remains one of the most reliable prognostic factors for decisions about treatment plans.

Due to the very heterogeneous nature of breast cancer, it is crucial to identify subgroups of patients with different risk factors and establish individual benefit/risk ratios in order to tailor breast cancer treatments and management.

1.1.3. Risk factors

1.1.3.1. Environmental risk factors

Similarly to most solid cancers, cases of breast cancer peak in the 60 to 64 age group, with age being one of the known risk factors. The second risk factor is gender, and relates to the female patient lifetime exposure to oestrogens (Kufe *et al*, 2003). Indeed breast cancer risk is influenced predominantly by gynaecological events in a woman's lifetime; an early menarche, a late first pregnancy, fewer births and a late menopause can all cause an increase in risk (Kelsey and Berkowitz, 1988). Oral contraceptives, hormone replacement therapy (HRT) and lifestyle factors, such as obesity and alcohol consumption, also have been identified as risk factors. For instance it has been established that 6% of UK female

breast cancer can be attributed to alcohol intake, a study on obesity has established that weight gain post-menopause of 22Ibs (9.9kg) leads to increase in risk of 18% and, finally, that the risk from passive smoking could lead to an increase in risk of up to 70% (Calle and Kaaks, 2004; Zhang *et al*, 2007; Kelsey and Berkowitz, 1988).

The changes in lifestyle observed in Western populations over the last 50 years have led to an increase in occurrence of most of these environmental risk factors, as well as a change in the approach to motherhood (fewer births, later in life), which might explain the increase in incidence observed in those countries (White, 1987; Parkin, 2011). Additionally, part of this increase has been caused by more cancer being detected by the implementation of routine mammographic screening. This increase in occurrence however has been associated with a decrease in mortality (Fig. 1.1A), attributed both to improvement in breast cancer treatments and the benefits of large scale screening, i.e. early detection.

1.1.3.2. Genetic risk factors

One of the strongest risk factors for developing breast cancer is family history of the disease. A woman's lifetime risk of developing breast cancer with no family history of the disease is estimated to be 8 to 10% but this risk can increase up to 87% in families with affected members. This risk is correlated with closeness of kinship with affected relatives (female or male), number of affected relatives, and age at onset of breast cancer in affected relatives (Eisen and Irwin, 2002). Interestingly, it was recently established that the relative risk of breast cancer for a woman with an affected brother is approximately 30% higher than for a female with an affected sister (Bevier *et al*, 2012).

BRCA mutations

BRCA1 and *BRCA2*, the two most important breast cancer susceptibility genes, were identified by linkage analysis and first cloned in the mid 1990s (Hall *et al*, 1990; Miki *et al*, 1994; Wooster *et al*, 1995). Approximately 20 to 25% of hereditary breast cancers are characterised by germline mutations in *BRCA1* (60 to 65% of cases) or *BRCA2* (35 to 40% of cases) with the prevalence of mutations

varying according to patient ethnicity, age and family history (Thompson and Easton, 2004). These mutations are rare in the general population but confer high risk of ovarian and breast cancer for the carriers (Antoniou *et al*, 2003), with a lifetime risk of 47 to 87% for breast cancer, together with an earlier onset of the disease (Fackenthal and Olopade, 2007). Genetic testing for mutations in these genes is now well established in high-risk families (Walsh *et al*, 2006).

The *BRCA1* gene encodes a nuclear phosphoprotein that plays a role in maintaining genomic stability, thus acting as a tumour suppressor. BRCA1 protein combines with other DNA damage proteins to form a large multi-subunit protein complex known as the BRCA1-associated genome surveillance complex (BASC) and plays a role in transcription, repair of double-stranded DNA breaks, and recombination (Wang *et al*, 2000). In patients with BRCA mutant breast cancer, a deficiency in DNA damage repair is observed. Inhibition of poly ADP ribose polymerase (PARP), which results in increased double stranded DNA damage specifically in cancer cells lacking BRCA, is a successful treatment strategy relying on the principle of synthetic lethality (Farmer *et al*, 2005; Fong *et al*, 2009).

Familial syndromes

The primary syndrome associated with the highest risk of breast cancer is hereditary breast and ovarian cancer syndrome, caused by mutations in the DNA repair genes *BRCA1* and *BRCA2*. However there are other hereditary cancer syndromes also associated with an increased risk of breast cancer and characterised by mutations uncommon in the general population. These include Li-Fraumeni syndrome, associated with germline mutations in *TP53* (Varley, 2003), Cowden disease, associated with germline mutation in *PTEN* (Marsh *et al*, 1999), Peutz-Jeghers syndrome, associated with truncation mutations in *LKB1* (Jenne *et al*, 1998). In addition to these genes, two other genes are associated with familial syndrome with a more moderate risk of breast cancer. Germline mutations in the *ATM* gene are found in patients suffering from ataxia-telangiectasia (de Jong *et al*, 2002). It has also been shown that a truncated variant of *CHEK2*, 1100delC, confers a two-fold increased relative risk for developing breast cancer (Meijers-Heijboer *et al*, 2002).

Low penetrance risk factors

Classical breast cancer susceptibility genes, such as *BRCA1* and *BRCA2*, have been estimated to account for only 25% of the familial breast cancer risk (Peto *et al*, 1999; Thompson and Easton, 2004; Antoniou and Easton, 2006). When including other known susceptibility genes and potential environmental factors, only a small portion of the familial cases of breast cancer are accounted for (Hopper and Carlin, 1992; Thompson and Easton, 2004), indicating that the majority of the risk factors remain undiscovered.

A new class of susceptibility genes or risk variants (discussed in more detail later in section 1.3), that confer a low disease risk to the individual but occur at high frequencies in the general population, was unearthed by advances in genomic analysis. These variants were identified with the emergence of large scale genome-wide association studies (GWAS), together with new statistical and bioinformatics tools. GWAS involve scanning a set number of markers (notably single nucleotide polymorphisms) across the complete genomes of two groups of individuals, patients with a certain disease (cases) and matched controls, in order to find genetic loci associated with the particular disease. The major advantage compared to linkage analysis was that high risk families were not required and the number of cases involved (thousands in each group) led to a greater power to discover risk factors (Mavaddat *et al*, 2010). Therefore we now see breast cancer as a polygenic disease, where a large contribution to the development of a tumour may be attributed to low penetrance factors such as particular polymorphisms. However the majority of the risk-associated SNPs occur in non coding regions of the genome complicating any attempt to investigate the function of those risk variants.

1.1.4. Breast cancer biomarkers and classification

Breast cancer is an heterogeneous disease, which encompasses a variety of distinct cellular abnormalities with distinct morphological features and clinical behaviours. Once diagnosed, carcinomas of the breast are described predominantly by their histological presentation and several biomarkers that have been found to be predictive of the treatment outcome, or patient survival.

Receptor status

Both oestrogen receptor alpha (ERα) and progesterone receptor (PR) have a prognostic value in breast cancer patients, although their ability to discriminate between low and high risk patients remains quite limited (Stewart *et al*, 1982). Patients with ER-positive tumours tend to have less aggressive disease with metastases targeting the bones and soft tissues (James *et al*, 2003), unlike ER negative patients, who tend to have earlier relapses and metastases in the liver, lung and central nervous system (Tham *et al*, 2006). However, differences between the two groups decrease as the disease progresses, with most ER positive tumours leading to ER negative metastases (Lower *et al*, 2005). ER positive tumours are often well differentiated and are associated with better prognosis, despite the fact that estradiol is a potent mitogen for receptor positive cells, as they can respond to endocrine therapy with Tamoxifen, an ERα antagonist (Heel *et al*, 1978). The best use of steroid hormone receptors is therefore not in the determination of prognosis but in the prediction of response to endocrine therapy and, therefore, the selection of optimal treatments.

Similarly the HER2/neu receptor is used in patient screening in order to determine the best treatment options. Amplification or overexpression of the *HER2* gene occurs in approximately 30% of early stage breast cancer and is associated with an unfavourable prognosis and high recurrence rate (Coussens *et al*, 1985; Yarden, 2001). However, HER2 positive breast cancer is now being treated successfully with several blocking antibodies (trastuzamab and pertuzamab) that target the HER2/neu receptor, in combination with chemotherapy, and these approaches have provided a five year increase in disease-free survival of patients (Romond *et al*, 2005).

Another subgroup of breast cancers are characterised as triple negative breast cancer as they lack active ERα, PR and HER2 receptors, and account for 10 to 17% of all breast carcinomas (Dent *et al*, 2007). These patients have the worst prognosis because of a lack of targeted therapy. Triple negative disease frequently affects younger patients (<50 years), is more prevalent in African-American women (Bauer *et al*, 2007) and is significantly more aggressive than other subtypes.

Molecular classification

Advances in molecular biology and high throughput technologies have permited the establishment of a new taxonomy for breast cancer based on gene expression profiles as examplified by the seminal Perou and Sorlie molecular classification of breast cancer (Perou *et al*, 2000; Sorlie *et al*, 2001). This classification regroups six molecular breast cancer subtypes : Luminal A, Luminal B, Luminal C, normal breast like, HER2 and basal-like. Some of these subgroups were, to some extent, already known by conventional histological classification and some degree of controversy about the reliability of this new classification has failed to generate internationally accepted definitions for some of these breast cancer groups, such as the basal-like group, which is generally defined as triple negative breast cancer (Badve *et al*, 2011).

1.1.5. Conclusion

Over 11, 500 women are expected to die of breast cancer in the UK this year (Data and Statistics, 2012). Breast cancer is still the second most common cancer related death among women in the UK, after lung cancer, despite several decades of research that have led to improvement in the screening, diagnosis and treatment of the disease. It has now been esablished that only 28% of the genetic risk factors for breast cancer are known (Michailidou K, 2013). This calls for a more thorough understanding of the genetics and molecular mechanisms that trigger and sustain cancer development, and will be key to the development of targeted therapies and the improvement of patient survival.

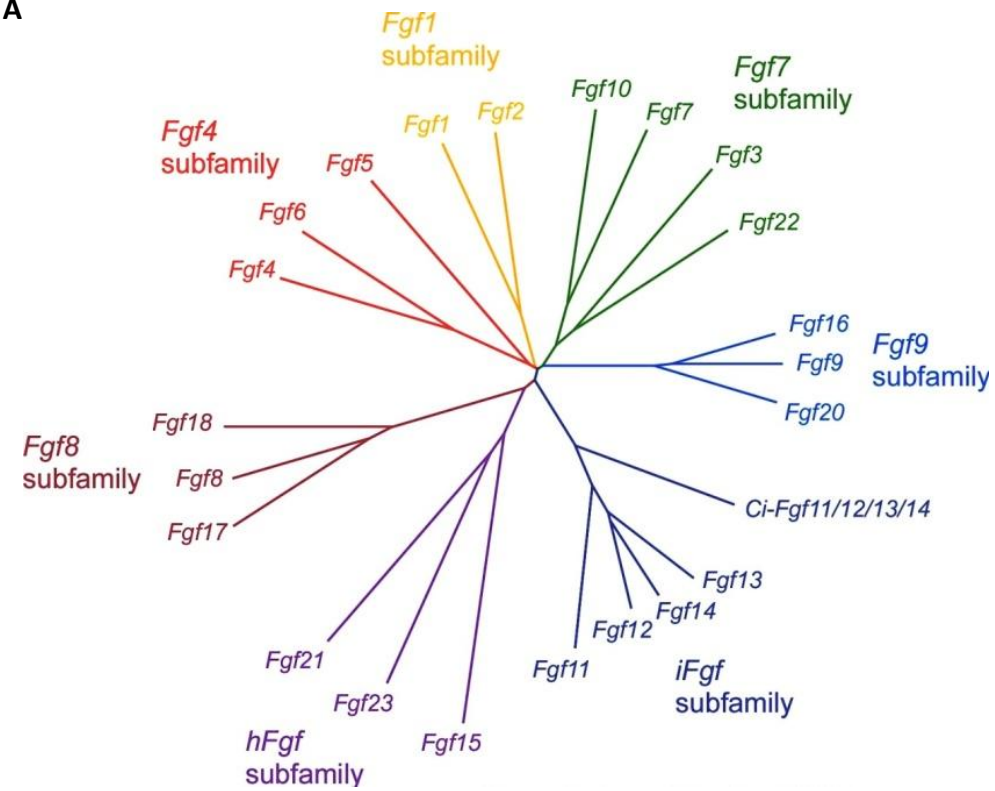
1.2. Fibroblast growth factors and their receptors

1.2.1. Fibroblast growth factors (FGFs)

The mammalian fibroblast growth factor (FGFs) family is composed of 22 members of intercellular and intracellular signalling molecules (from 17 to 34 kDa), 18 of which are functional ligands (FGF1 to 10 and FGF16 to 23) for the fibroblast growth factor receptors (FGFRs). They can be grouped according to their sequence homology and receptor binding affinity into seven different FGF subfamilies (Fig. 1.2A) (Itoh and Ornitz, 2008). FGF1 is the most promiscuous ligand that can bind to multiple receptors. This promiscuity in binding is in contrast to specific ligands, like FGF7 for instance, which is specific for FGFR2b (Fig. 1.2B) (Ornitz *et al*, 1996).

The structure and function of the FGF family of ligands are conserved throughout vertebrate evolution, with 13 to 71% amino acid identity (Ornitz and Itoh, 2001), and they are found in a wide range of organisms from nematodes to humans. The large number of FGFs is thought to have evolved through phases of global gene duplication in the period before the emergence of vertebrates (Coulier *et al*, 1997). Most FGFs share an internal region of similarity, with 28 highly conserved and 6 identical amino acids. Not surprisingly, ten of these residues are responsible for interactions with their receptor (Plotnikov *et al*, 2000). Most FGFs, with the exception of FGF1 and FGF2 (Mignatti *et al*, 1992), are secreted proteins and include amino-terminal signal peptides for trafficking via the endoplasmic reticulum and the Golgi apparatus (Ornitz and Itoh, 2001). In addition, FGFs can bind heparin and heparan sulphate-proteoglycans (HSPG) present at the cell surface and in the extracellular matrix (Fig. 1.3C). HSPGs behave as low affinity receptors for the FGFs and are believed to protect them against degradation by proteases or thermal denaturation (Copeland *et al*, 1991) but also to facilitate the assembly and activation of the FGFR/FGF complex by physically anchoring them in close proximity at the plasma membrane (Rapraeger *et al*, 1991). In cell culture, FGFs can stimulate cell growth, migration and differentiation (Eswarakumar *et al*, 2005). *In vivo*, they are responsible for many different cellular functions, principally during development. FGFs are widely expressed but follow a strict spatial and temporal pattern in the embryo, where they are crucial for development and

A



B

	Isoforms	Ligands
FGFR2	b	FGF1, FGF3, FGF7, FGF10, FGF22
	c	FGF1, FGF2, FGF4, FGF5, FGF6, FGF8, FGF9, FGF16, FGF17, FGF18, FGF20

Figure 1.2: The fibroblast growth factor family

A) Phylogenetic tree of the fibroblast growth factors showing seven subgroups of closely related peptides (adapted from (Itoh and Ornitz, 2008)). B) FGFR2 isoforms and ligand specificity.

differentiation of highly hierarchical organs such as the skeleton, lungs, and the circulatory and central nervous systems (Ciruna and Rossant, 2001). They also play important roles in the adult organism in wound healing, tissue repair, angiogenesis and as homeostatic factors (Turner and Grose, 2010).

1.2.2. Fibroblast growth factor receptors (FGFRs)

Fibroblast growth factor receptor genes (*FGFRs*) encode transmembrane receptor tyrosine kinases (RTKs). They contain two or three immunoglobulin-like domains and a heparin-binding domain in their extracellular portion (Fig. 1.3A). Alternative mRNA splicing of the second half of the third immunoglobulin-like domain in FGFR1-3 gives rise to various FGFR isoforms that differ in their ligand binding affinities (Fig. 1.3B) (Zhang *et al*, 2006). This alternative splicing event is mostly tissue and cell type specific: the IIIb isoforms are more commonly expressed on cells from epithelial lineage and the IIIc isoform, in the mesenchymal lineage (Orr-Urtreger *et al*, 1993). In conjunction, expression of their specific ligands occurs in adjacent tissues, leading to directional paracrine signalling between epithelial and mesenchymal cells. A fifth related receptor, FGFR5 (also known as FGFRL1), can bind FGFs but lacks a tyrosine kinase domain, thus potentially acting as a negative regulator of FGF signalling (Steinberg *et al*, 2009). Although the different FGF receptors have overlapping expression patterns and functional similarities, they also mediate very specific effects depending on cellular differentiation and context (Dailey *et al*, 2005).

1.2.2.1. Signalling

Upon interaction with a ligand, stabilised by the cell surface HSPGs and dimeric Grb2 (Lin *et al*, 2012), a conformational shift in receptor dimer structure elicits transphosphorylation of the intracellular kinase domains (Furdui *et al*, 2006; Mohammadi *et al*, 1996). The phosphorylated tyrosine residues of the cytoplasmic portion of the receptor can then act as a docking site for adaptor proteins containing Src homology-2 (SH2) and phosphotyrosine-binding (PTB) domains (Fig. 1.3C) (Mohammadi *et al*, 1996). FGFRs signal through four pathways that regulate cell proliferation, survival and differentiation: the MAPK, the PI3K/AKT, the

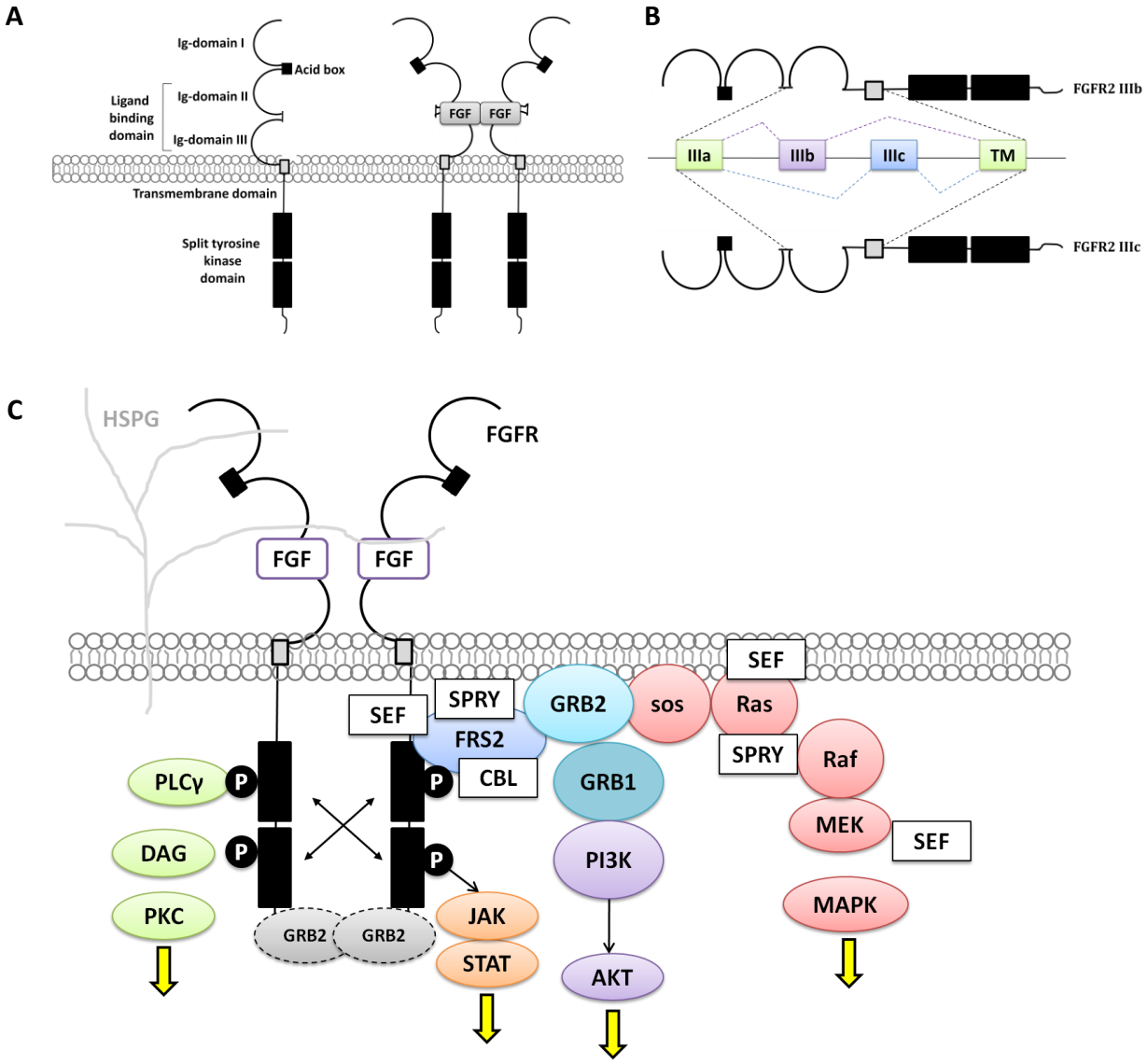


Figure 1.3: FGF receptor structure and organisation at the plasma membrane

A) The basic structure of the FGF–FGFR complex comprises two receptor molecules, two FGFs and one heparan sulphate proteoglycan chain. The FGFR consist of three extracellular immunoglobulin (Ig) domains, a single transmembrane domain and an intracellular split tyrosine kinase domain. The second and third Ig domains form the ligand-binding pocket and have distinct domains that bind both FGFs and HSPGs. B) Ligand-binding specificity is generated by alternative splicing of the Ig III domain. The first half of Ig III is encoded by an invariant exon (IIIa), which is spliced to either exon IIIb or IIIc, both of which splice to the exon that encodes the transmembrane (TM) region. C) FGFR signals through four pathways, the MAPK pathway (red), the PI3K/AKT pathway (purple) the JAK/STAT pathway (orange) and PLCγ pathway (green). Both MAPK and AKT signalling require the presence of an FGFR specific adaptor protein: FRS2. Proteins in black boxes are inhibitors of the FGFR signalling pathways (Adapted from Dickson *et al*, 2000).

JAK/STAT and the PLCγ pathways (Fig. 1.3C). The predominant signalling pathway activated downstream of FGFRs in development is MAPK signalling (Corson *et al*, 2003), which promotes cell proliferation.

FRS2, a major adaptor that links FGFRs to the ERK and PI3K/AKT pathways, binds to the active kinase domain of the receptor and becomes phosphorylated on specific tyrosine residues (Kouhara *et al*, 1997; Ong *et al*, 2001). GRB2, another adaptor protein, is then recruited and can transduce the signal to ERK and PI3K/AKT pathways. Activation of the AKT pathway results in anti-apoptotic signalling, as well as cell growth and proliferation cues (Gotoh, 2008), whereas MAPK activates transcription factors involved in control of the cell cycle. Indeed, it has been shown that cyclin D1 and D2, master regulators of the cell cycle, are downregulated when FGFR signalling via MAPK is inhibited (Koziczak *et al*, 2004).

Independently of adaptor proteins, the active receptor can bind to and phosphorylate PLCγ (Mohammadi *et al*, 1991) which leads to the activation of PKC, via DAG, and therefore reinforce the MAPK pathway.

The JAK/STAT pathway is also independent of adaptor proteins and leads to the translocation of STAT to the nucleus, which in turn directs the transcription of target genes associated with proliferation, differentiation and apoptosis (Darnell, 1997).

FGFR signalling can differ between receptors in nature and strength of the signal produced (i.e. the pathway activated). For instance, it was found that FGFR4 produced a weaker signal than either FGFR1 or FGFR2, particularly with respect to responses involving PLCγ and FRS2, whereas FGFR3 produces signals similar to FGFR1 (Raffioni et al, 1999). Internalisation and degradation of the receptor constitute an important aspect of FGFR regulation and are mediated by key FGF negative regulators such as CBL (responsible for FGFR and FRS2 ubiquitination) (Wong *et al*, 2002), SPRY (which interrupts the GRB2/FRS2 complex and prevents Ras phosphorylation) and SEF (inhibits FGFR and FRS2 phosphorylation, prevents phosphorylated ERK from migrating to the nucleus) (Fig 1.3C).

1.2.3. FGFs in the mammary gland

Normal development of the mammary gland relies on highly orchestrated interactions between epithelial and mesenchymal cells that start during embryogenesis and continue during puberty where the glands undergo growth and differentiation. It has been studied most closely in the mouse (Fig. 1.4), nevertheless the basic processes and signalling mechanisms hold true for human development (Hynes and Watson, 2010).

Bulb-like structures on the tips of the epithelial ducts, called terminal end buds (TEBs), proliferate and penetrate into the fat pad as the ducts elongate (Fig. 1.4). TEBs diverge and secondary branches appear, until the entire fat pad is filled with a network of branched ducts. This simple epithelial tree within the mammary fat pad remains quiescent until the onset of puberty, when ovarian steroid hormone production commences. During repeated oestrous cycles, the ductal network increases in complexity and side-branches grow under the control of progesterone. In response to prolactin, alveolar structures bud off the ductal system during pregnancy, and these differentiate into milk-producing sacs. Steroid hormones, growth hormone and prolactin are the master regulators of mammary growth and pregnancy-induced differentiation, whereas epidermal growth factor (EGF) and FGFs have more specific and specialised roles (Hynes and Watson, 2010).

Alternative splicing of FGFRs in the third IgG loop yields different isoforms of the receptors that bind different FGFs (Fig. 1.3B) and are expressed in epithelial or mesenchymal compartments, respectively. FGF10 and its main receptor FGFR2-IIIb have important roles during embryonic mammary gland development. Starting at embryonic day 10.5 in mice, FGF10 acts on FGFR2-IIIb in the ectoderm to initiate induction and positioning of the future mammary glands. The ligand and the receptor are both essential for induction of placode pairs (Mailleux *et al*, 2002; Veltmaat *et al*, 2006).

During ductal growth, multiple FGFs (FGF1, FGF2, FGF4, FGF7, and FGF10) as well as FGFR1 and FGFR2 are expressed (Schwertfeger, 2009). It has been shown that the epithelial specific FGFR2-IIIb isoform is responsible for ductal outgrowth, and TEBs usually express high levels of FGFR2 (Parsa *et al*, 2008). Indeed, it was

demonstrated that *FGFR2*-null glands penetrate the fat pad more slowly and show fewer branch points compared with wild-type controls (Lu *et al*, 2008). Finally, some FGFs are expressed during pregnancy and lactation (Coleman-Krnacik and Rosen, 1994), their role remains unclear however.

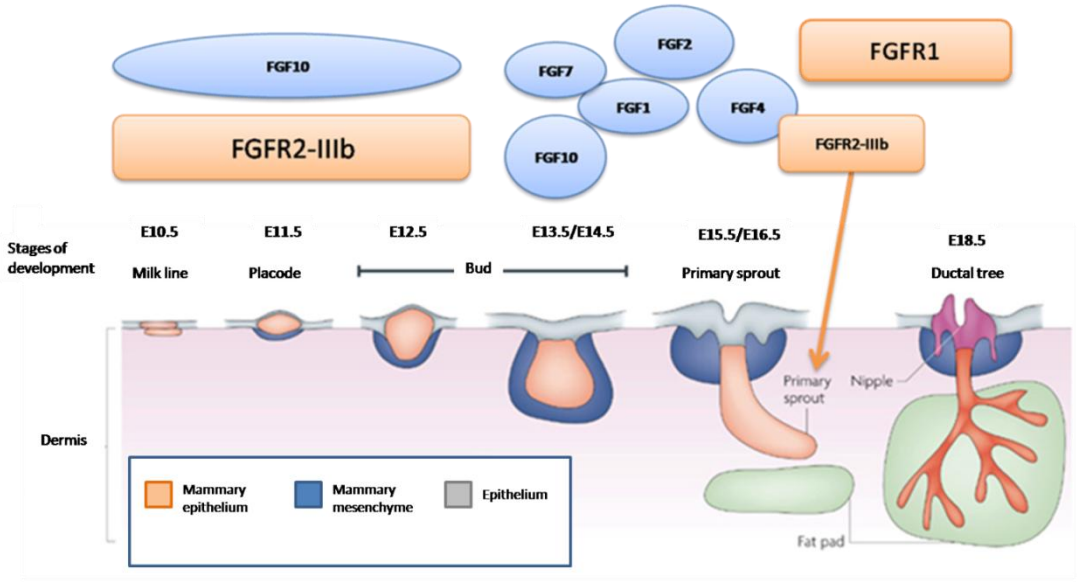


Figure 1.4: Morphological stages in the embryonic development of the mouse mammary gland and FGF signalling

Around embryonic day 10 (E10) of mouse development the milk line (orange) is defined by a slight thickening and stratification of the ectoderm (grey) as shown here in this series of cross sections through the trunk. On E11.5 the milk line breaks up into individual placodes (orange) and the underlying mammary mesenchyme (blue) starts to condense. Over the following days the placodes sink deeper into the dermis and the mammary mesenchyme becomes organised in concentric layers around the mammary bud (orange). Starting on E15.5, the mammary epithelium (orange) starts to proliferate at the tip and the primary sprout pushes through the mammary mesenchyme towards the fat pad (green). On E18.5 the elongating duct has grown into the fat pad and has branched into a small ductal system. The cells of the mammary mesenchyme have formed the nipple, which is made of specialised epidermal cells (purple)(Robinson, 2007).

1.2.4. *FGFR2* and cancer

Germ line and somatic FGFR mutations are known to play a role in a range of diseases, most notably skeletal disorders and cancers; they are predominantly activating mutations resulting in sustained signalling promoting survival, proliferation and angiogenesis, therefore indicating that *FGFR2* can act as an oncogene (Grose and Dickson, 2005). However, *FGFR2* can be ambivalent, and there are certain types of cancers, like bladder or skin, in which *FGFR2* has been associated with a tumour suppressive phenotype (Gartside *et al*, 2009). Generally, activating mutations in *FGFR2* are found most frequently in endometrial cancer (Byron *et al*, 2008) and are rare in breast cancer (Greenman *et al*, 2007). Mouse models of mammary carcinogenesis have long established the FGF signalling pathway as a major contributor to tumourigenesis (Grose and Dickson, 2005), and a mouse mammary tumour virus (MMTV) insertional mutagenesis screen for genes involved in breast cancer identified *FGFR2* and *FGF10,* one of its ligands, as potential oncogenes in the breast (Theodorou *et al*, 2007).

Although *FGFR2* amplification has been reported in several publications (Heiskanen *et al*, 2001), a genome-wide copy number variation screen found *FGFR2* to be amplified in only 2 out of 161 primary breast cancer samples (1.2%) (Kadota *et al*, 2009). This result was confirmed in other similar screens looking at a large group of unselected breast cancer samples (Adelaide *et al*, 2007), and sometimes amplification or losses of *FGFR2* were not observed at all (Andre *et al*, 2009). When looking at specific types of cancers like triple negative disease, amplification was more common, reaching 4% of cases and was not found in other subtypes. A concomitant increase in FGFR2 mRNA levels was also observed in those samples, suggesting a potential role for FGFR2 in triple-negative breast cancer (Turner *et al*, 2010). FGFR2 expression was also significantly higher in familial breast cancer patients with germ line mutations in BRCA1 and BRCA2 (Bane *et al*, 2009). Functional studies in cell lines have implicated *FGFR2* as playing a role in tumourigenesis, with an alternative splicing in the C-terminal domain of FGFR2 giving rise to a more transforming isoform (Tannheimer *et al*, 2000). The presence of this transforming isoform in patients is, however, quite rare.

Additionally, high FGFR2 expression has been associated with poor prognosis and lower (disease-free) survival rates (Sun *et al*, 2012). The exact molecular mechanism linking FGFR2 and breast cancer is however not clear and remains to be determined. Finally, *FGFR2* was also identified as a new risk locus for ER positive breast cancer in two independent GWAS studies (Easton *et al*, 2007; Hunter *et al*, 2007), and this novel mechanism is the focus of this study.

1.3. Single nucleotide polymorphisms

Deleterious somatic and germinal mutations are rare but play a determinant role in the emergence of cancer and other Mendelian disorders. Common and frequent genetic variations (polymorphisms) on the other hand may also play a role in cancer susceptibility but are more complex to identify.

The most common type of genetic polymorphisms are single nucleotide polymorphisms (SNPs), accounting for 90% of human DNA variations (Collins *et al*, 1998). In one of the first comprehensive reviews on the subject, Brookes describes SNPs as "single base pair positions in genomic DNA at which different sequence alternatives (alleles) exist in normal individuals in some population(s), wherein the least frequent allele has an abundance of 1% or greater" (Brookes, 1999). Subsequently, the human genome project identified more than 1.42 million SNPs, which corresponds, on average, to one SNP every 1,300 bp (Taillon-Miller *et al*, 1998; Lander *et al*, 2001; Sachidanandam *et al*, 2001). However, SNPs are not distributed uniformly over the entire human genome and their distribution sheds light on the unique properties and history of each genomic region.

Some SNPs have a well defined impact on phenotypes, as they are located in the coding region of genes, and create a non-synonymous substitution that leads to amino-acid change or, in the 3'-UTR region of messenger RNA, affects the binding of micro RNAs (Abelson *et al*, 2005). However, the rate of nucleotide difference is, understandably, four-fold lower within coding exons compared to non-coding regions and only half of those changes result in non-synonymous codon-change (Li and Sadler, 1991; Nickerson *et al*, 1998). The importance of SNPs located in non coding regions is now recognised but their functions remain elusive.

The Functional Single Nucleotide Polymorphism (F-SNP) database was created in 2007 to integrate information about the functional effects of SNPs (Lee and Shatkay, 2008). The aim was to predict the effect of SNPs at the splicing, transcriptional, translational and post-translational level using bioinformatic tools. Their results show that an estimated 20% of SNPs disrupt genomic regions known to be functional, including splice sites and transcriptional regulatory regions. Recently, the first publications of the ENCODE consortium (Consortium *et al*, 2012;

Sanyal *et al*, 2012; Thurman *et al*, 2012) have shed more light on the impact of functional information within the non coding regions of the genome and the importance of variation in regulating gene function.

1.3.1. Functional SNPs in cancer

The characterisation of human SNPs and their role in phenotypes remains a challenge and the study of rare genetic variants is only becoming possible in recent years with advances in data mining and new genetic technologies. Although it is now possible to sequence the entire human genome, this information alone is not sufficient to understand the biological implication of most sequence variations. SNPs have been used mostly by research groups as genomic markers to identify regions that are associated with disease. Within a single chromosome, a conserved combination of SNPs can be concentrated at a specific region, usually implying a region of medical or research interest. For instance, an extreme example is the human leucocyte antigen (HLA) region on chromosome 6, in which a high SNP density is observed (5 to 10% of nucleotide diversity), reflecting the fact that a diverse combination of HLA alleles has been maintained for many thousands of years by natural selection, since they emerged long before the divergence of *Homo sapiens* from its common ancestor (Hughes *et al*, 2005).

Similarly to somatic mutations, SNPs located in coding regions are known to play a role in cancer progression or response to treatment. In the FGFR family for example, a functional FGFR4 polymorphism (Gly388Arg) located in the transmembrane domain of the receptor was first found associated with colorectal cancer (Spinola *et al*, 2005) and more recently, was used as a predictive marker for outcome of treatment with an mTOR inhibitor in pancreatic neuroendocrine tumours (Serra *et al*, 2012).

Methods to study non-exonic SNP functions rely on the study of expression level of genes and prediction of binding sites for transcription factors. To date, only a few regulatory SNPs have been characterised at the genetic and phenotypic level. Members of the pathway of the tumour suppressor p53, which is involved in at least half of human cancers, provide one such example. *MDM2 is* a direct negative regulator of p53 and SNP309, located in the *MDM2* promoter, was found to

increase the binding affinity of the transcription factor Sp1 which leads to overexpression of MDM2 and, subsequently, to an attenuation of the p53 pathway. Patients with hereditary (Li Fraumeni) and sporadic cancers (soft tissue sarcoma) that are homozygous for the G allele of SNP309 show an accelerated progression of the disease (Bond *et al*, 2004; Bond *et al*, 2006).

The methods usually applied in the study of non-intronic SNPs involve indirect assessment of transcription factor binding by electrophoretic mobility shift assay (EMSA), chromatin immunoprecipitation (ChIP) and reporter assay. Recently, whole genome screens are emerging, analysing the total number of binding sites of certain transcription factors and their relation with known polymorphisms (Spivakov *et al*, 2012). But these whole genome approaches are more appropriate to identify general patterns in transcription factor dynamics than to shed light on individual functional SNP. New approaches are therefore needed in order to study the impact of such polymorphisms on cell behaviour.

1.3.2. The *FGFR2* haplotype

Breast cancer was one of the first diseases to be studied using the new power of GWAS and many new susceptibility genes were subsequently identified. In 2007, two studies identified a region of the second intron of the *FGFR2* gene as the most significant locus associated with increased risk for sporadic postmenopausal ER positive breast cancer in patients of European descent (Easton *et al*, 2007; Hunter *et al*, 2007). This haplotype lies within a 25 kb linkage disequilibrium block almost entirely within intron 2 of *FGFR2* (Fig. 1.5A). This particular region was found to be conserved among mammals, showed Histone 3 acetylation marks and DNaseI hypersensitivity clusters, all of which are often found near active regulatory elements (Fig. 1.5B). The carriers of the risk allele of the most significantly associated SNP, rs2981582, showed a risk of breast cancer 1.26 times greater than non-carriers (Fig. 1.5C). The haplotype was not in linkage disequilibrium with other SNPs elsewhere in the coding region of the gene and its intronic location

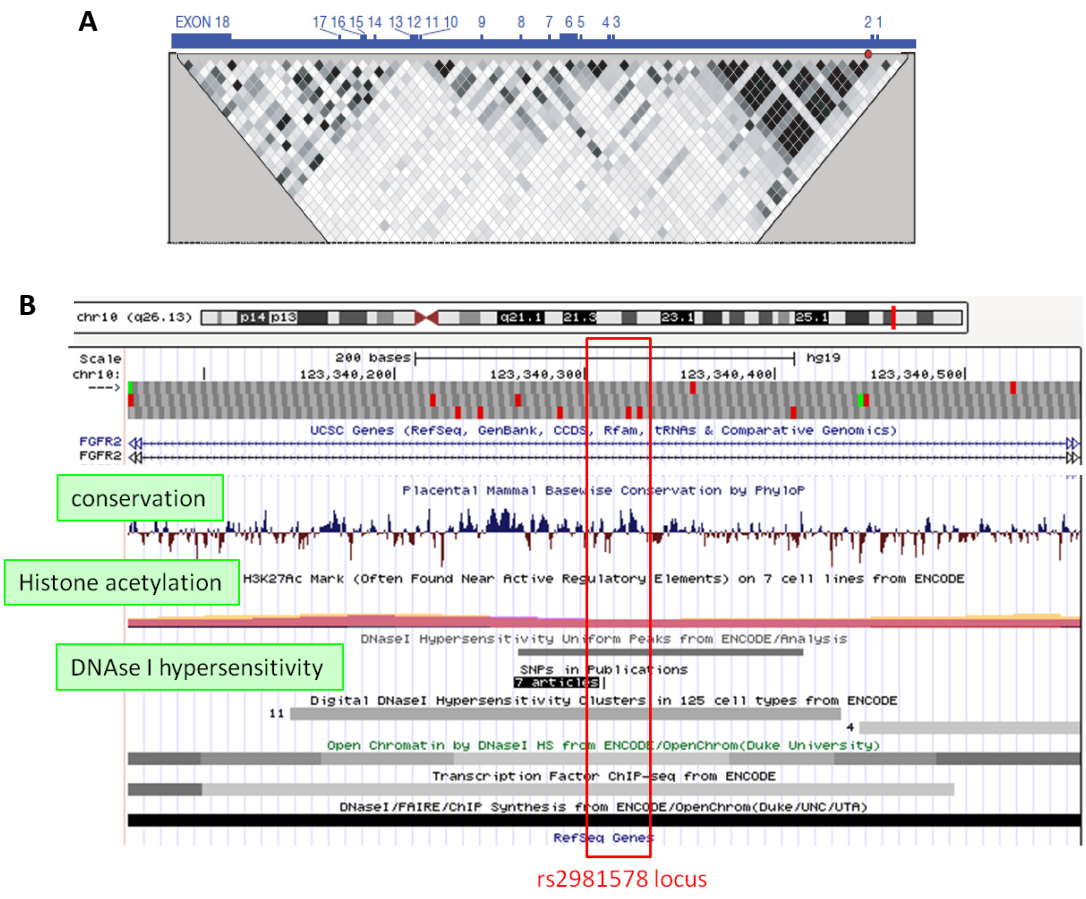


Figure 1.5: The *FGFR2* locus

A) Map of *FGFR2* showing the linkage disequilibrium block in the large second intron. The gene is 126 kb long and in reverse 3'-5' orientation on chromosome 10 (Easton *et al*, 2007). B) Genomic landscape around *FGFR2* haplotype region containing rs2981578 (red box), showing mammalian sequence conservation, H3 histone acetylation marks, DNaseI hypersensitivity regions on the UCSC genome Browser (GRCh37/hg19). C) Estimated risk increase in developing breast cancer associated with the different alleles of rs2981578 (Easton *et al*, 2007; Hunter *et al*, 2007).

indicated that its association with breast cancer most likely was mediated through regulation of *FGFR2* expression.

rs2981578, a putative functional SNP found within this haplotype, mapped to binding sites for the transcription factors Oct1/Runx2 (Meyer *et al*, 2008). Using ChIP, it was found that the disease-associated allele (G) created a new binding site for the Runx2 transcription factor. The disease associated allele (G) is also the ancestral allele and is found in the vast majority of individuals of African origin and to a lesser extent in other populations (Fig. 1.6). Luciferase assays, using a cloned region of intron 2, revealed that the presence of several Runx2 and an Oct1 binding sites in close proximity caused an increase in luciferase expression, thus suggesting a molecular explanation for the risk phenotype. Indeed, the Runx2/Oct1 complex has also been identified on the promoter of *β-casein*, a mammary gland specific gene. Paul Shore's group used ChIP, RNA interference and promoter mutagenesis to show that the complex acted as an enhancer of *β-casein* expression (Inman *et al*, 2005). Moreover, the increased risk conferred by the minor *FGFR2* allele associated most strongly with ERα positive breast tumours, suggesting a potential interaction between the two pathways (Easton *et al*, 2007; Hunter *et al*, 2007; Udler *et al*, 2009). The other risk allele looked at in this study was rs7895676, which displayed a reduced binding capacity to C/EBPβ (Meyer *et al*, 2008). In addition, a meta-analysis found that *FGFR2* rs2981582 was significantly associated with the risk in BRCA2 mutation carriers predominantly (Antoniou *et al*, 2008). Indeed, the vast majority of BRCA1 breast cancer tumours are oestrogen receptor negative whereas BRCA2 tumours have an ERα status distribution similar to that of unselected breast cancers, of which the majority are ER positive. However BRCA1 patients can sometimes present with an ER positive cancer and the same association was found in this particular group (Mulligan *et al*, 2011) indicating a strong link between the oestrogen receptor and the *FGFR2* risk alleles, via a mechanism that remains to be elucidated.

Several other successive GWAS have confirmed the association of the *FGFR2* risk alleles with breast cancer. Different populations were tested: Jewish and Israeli (Raskin *et al*, 2008), Tunisians (Shan *et al*, 2012), Hispanic (Slattery *et al*, 2011),

rs2981578	G	A
ALL	0.623	0.377
AFR	0.921	0.079
AMR	0.547	0.453
ASN	0.535	0.465
EUR	0.532	0.468

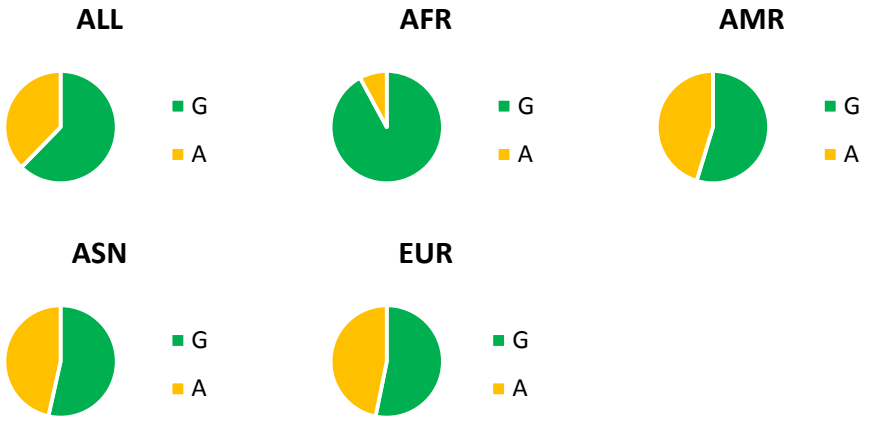


Figure 1.6: 1000 Genomes population data for rs2981578 allele frequencies

Allele frequencies for rs2981578 in the populations of the 1000 Genomes project. G is the ancestral allele and the risk allele for breast cancer (green), A is not associated with risk (yellow). Super population codes that regroup data from several populations: AFR: African, AMR: Ad Mixed American, ASN: East Asian, EUR: European. When the code ALL is used this means that all individuals from that data set are being considered (1000 Genomes Project, 2008-2012; Genomes Project, 2010)

African American (Udler *et al*, 2009; Zheng *et al*, 2009; Barnholtz-Sloan *et al*, 2010; Barnholtz-Sloan *et al*, 2011), Chinese (Long *et al*, 2010; Chan *et al*, 2012), Korean (Han *et al*, 2011), and all showed association with breast cancer and *FGFR2* haplotype.

In one study, only a trend towards association was found in African American women (Hutter *et al*, 2011). Additionally, male breast cancer was also found to be associated with the *FGFR2* haplotype (Orr *et al*, 2011), although the same association did not reach significance in another study (Orr *et al*, 2012). This result also strengthens the hypothesis of a cross-talk between FGFR2 and ERα as male breast cancers are almost entirely ER positive.

1.4. Zinc finger nucleases: targeted genome editing

Existing methods for targeted gene modification *in vitro* require several rounds of homologous recombination and drug selection to isolate rare desired events - a laborious and time consuming process that has limited the access to certain cell models. Zinc-finger nucleases (ZFNs) have become powerful tools for gene manipulation and offer a potential solution to this problem.

ZFNs are engineered to introduce a double-strand break (DSB) at any user-defined genomic locus. The double-strand break is resolved by the cell's own DNA repair machinery, while introducing modifications of the original sequence. This new technology is very promising thanks to its high specificity and the many potential applications that have arisen in recent years.

1.4.1. FokI restriction enzyme

ZFNs are synthetic modular molecules made from the fusion of zinc-finger DNA-binding domains to the catalytic domain of the endonuclease *FokI* (Fig. 1.7A) (Smith *et al*, 2000; Bibikova *et al*, 2001; Mani *et al*, 2005). Natural *FokI* is a two-domain protein (heterodimer) that binds a specific 5 bp DNA sequence and cuts at a distance away on the two strands. Chandrasegaran and colleagues were the first to show that the Type IIS restriction endonuclease *FokI* is a two-domain protein, with separable DNA recognition and cleavage functions. Change of the original cleavage target site was achieved by replacing the natural DNA-binding domain of *FokI* with one with different recognition binding site specificity (Kim *et al*, 1996). In ZFNs, the site of DNA cleavage is therefore determined by the recognition specificity of modular zinc fingers subunits.

Doyon *et al* identified critical residues involved in dimerisation, and used these residues to engineer ZFNs that have superior cleavage activity, while preventing homodimerisation (Fig. 1.7B) (Doyon *et al*, 2011). Two ZFNs are required, since the ZFN endonuclease activity is dependent upon dimerisation of heterogeneous *FokI* subunits. The prevention of homodimer formation in the new generation ZFNs has contributed greatly to reduced toxicity (Miller *et al*, 2007).

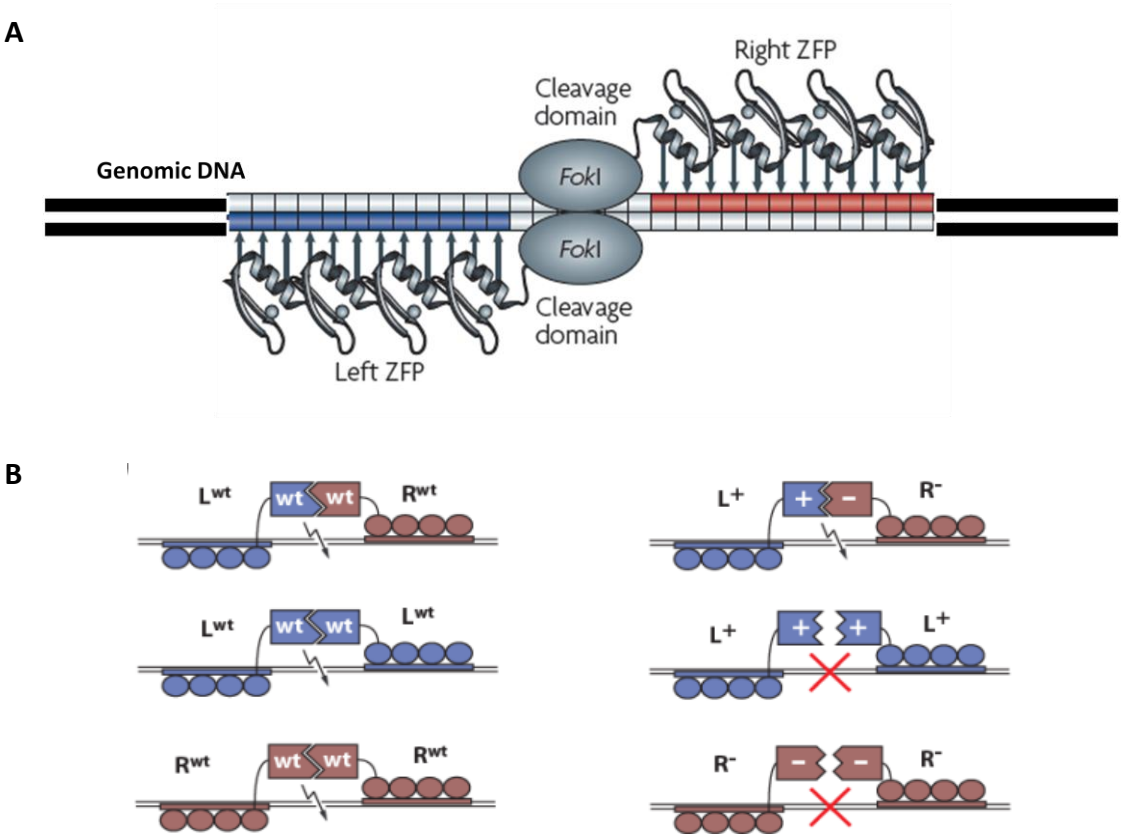


Figure 1.7: Spatial organisation of Zinc Finger Nucleases at the target site

A) Each ZFN consists of two functional domains: the zinc finger protein (ZFP), a DNA-binding domain comprised of a chain of four zinc finger domains, each recognising an unique sequence of DNA (left and right ZFP), and DNA cleavage domain formed of the two endonuclease domains of *FokI*, working as a highly-specific pair of 'genomic scissors'. B) New *FokI* protein design with obligate heterodimers (L^{+} and R^{-}, right), as opposed to potential homodimers (L^{wt} and R^{wt}, left), in order to eliminate off-target activity by homodimer species (Miller *et al*, 2007).

1.4.1. Zinc Finger proteins

The Cys2His2 zinc finger is a module of about 28-30 amino acids that is commonly found in transcription factors (Razin *et al*, 2012). ZFNs consist of an α-helix, two β-sheets, and a single zinc atom. Each finger recognises a particular triplet of DNA bases through contacts in the DNA major groove, many of which occur naturally. They can however be rearranged easily by changing the identities of the residues that alter the DNA recognition specificity of the finger to recognise all the potential DNA triplets. This ability to modify the DNA binding specificity of these molecules allows targeting of virtually any desired site in the genome.

1.4.2. Genome editing

ZFN-induced double strand breaks can stimulate two different DNA repair pathways: homologous recombination (HR) (Moynahan and Jasin, 2010) or non-homologous end joining (NHEJ) (Lieber, 2010). The latter is very inaccurate and, unlike homologous recombination, does not rely on sister chromatid homology. The homologous recombination pathway has evolved to deal with stalled DNA replication forks and double-strand breaks (endogenous and exogenous) and also contributes to genetic recombination (Moynahan and Jasin, 2010). However, in the absence of DNA damage, these recombination events have a very low efficiency. ZFNs enhance this efficiency by creating a double-strand break in a specific locus, therefore directing the DNA repair machinery to this site (Urnov *et al*, 2005). A repair template (plasmid DNA that contains the targeted genetic change) can be used for genome editing, replacing the sister chromatid, leading to targeted sequence replacement. Indeed, the highly homologous repair matrix is transfected in the cell at a high concentration, shifting the balance towards modification rather than normal repair. The repair template can introduce deletions, substitutions or insertions at the target site, along with selector genes that may be used for screening. Several kinds of DNA template have been tested with variable efficacy.

A recent study using ZFNs for gene knockout showed that the modification frequency of the *Glutamine Synthetase* (*GS*) gene reached 25% in CHO cells (Liu *et al*, 2010), which is much greater than the average 1% rate normally achieved with classic homologous recombination techniques. Indeed, in mammalian cells the

targeted DNA predominantly integrates in a random fashion rather than through homologous recombination, and it was estimated that for every one gene targeting event, there will be 10 to 20,000 random integrations (Sedivy and Sharp, 1989; Porteus and Baltimore, 2003).

The possibility of creating double strand breaks with high efficiency and at specific sites in the genome forms the basis for a wide range of applications for therapies and research. This technology has first been used in models like *Drosophila melanogaster* (Bibikova *et al*, 2002), *Arabidopsis thaliana* (Lloyd *et al*, 2005), zebrafish (Doyon *et al*, 2008) and rats (Geurts *et al*, 2009). For instance, ZFNs have been used successfully to create knockout mammalian cell lines (Liu *et al*, 2010), but this method can also be applied to rodent embryonic stem cells (Geurts *et al*, 2009) and constitutes a novel and rapid method to create new knockin or knockout mice. As an alternative to the use of stem cells, direct embryo injection of ZFN-encoding mRNA has been used in the fruit fly and zebrafish to generate heritable knockout mutations at specific loci (Carroll, 2008). Multiple pairs of ZFNs can also be used to remove large segments of genomic sequence (Lee *et al*, 2009).

Moreover, clinical applications of ZFN-gene therapy open new possibilities: dysfunctional genes (with a known, limited mutation) could be repaired directly in patients using a ZFN-based method. ZFNs can be used to disable dominant mutations in heterozygous individuals by producing double strand breaks in a mutant allele which will, in the absence of an homologous template, be repaired by non-homologous end-joining (NHEJ). The error-prone repair will result in deletion or insertion of base-pairs, producing a shifting in the reading frame and preventing the production of the mutated protein. The private company Sangamo Bioscience is currently testing (ongoing Phase 1/2 clinical trial) a ZFN mediated genome-editing of T-cells for the treatment of HIV/AIDS by targeting the receptor CCR5 (Holt *et al*, 2010; Ledford, 2011). CCR5 is a co-receptor for HIV entry into T-cells and, if CCR5 is not expressed on their surface, HIV infects them with lower efficiency. Naturally occurring mutations of CCR5 have been identified in individuals resistant to infection with the most common strain of HIV (Liu *et al*, 1996). The mutation CCR5delta32 leads to the expression of a truncated, and non-functional CCR5 protein, with no other observable deleterious effect (Perez *et al*,

2008; Holt *et al*, 2010; Moehle *et al*, 2007; Urnov *et al*, 2005). Their ultimate aim is to edit the immune cells of the infected patient, and then re-implant them in that same patient. Another therapeutic approach is to target episomal viral DNA. For instance, promising results have been obtained in the treatment of Hepatitis B, where the ZFN targeted cleavage is directed toward HPV DNA (Cradick *et al*, 2010).

1.4.3. Conclusion

Although the first ZFN was reported more than 10 years ago, the number of publications in this field has increased remarkably in the past few years and improved ZFNs are now more commonly available to researchers.

The capability of interfering with and manipulating gene sequences is one of the most reliable ways of learning about the importance of that sequence, especially when looking at non coding DNA sequences. In the study of risk-alleles associated with diseases, ZFNs show therefore great potential for basic research.

1.5. Aims and Objectives

The second intron of *FGFR2* contains SNPs that are associated with an increased risk of developing breast cancer. Notably rs2981578 has been identified as a putative functional SNP, modulating the binding of transcription factors.

The aim of this project was to study the effect of rs2981578 on *FGFR2* expression. The first objective was to create new cell line models that differ in this particular SNP status. The *in vitro* system created consisted of multiple isogenic mammary epithelial cell lines that represented the different allelic versions of rs2981578.

The second objective was to use these cells in functional studies to identify the mechanism by which the *FGFR2* SNP variant confers an elevated risk for the development of breast cancer.

The third objective was to assess the level of allele specific expression of *FGFR2* in a cohort of breast cancer patients with ER positive tumours.

The final, distinct but related objective was to use the *FGFR* ZFNs to allow insertion of a replacement cDNA into the endogenous *FGFR2* locus in order to induce specific FGFR2 mutations and assess the effect of such mutant proteins, expressed at physiological levels, on both receptor and cell behaviour.

CHAPTER 2

MATERIALS AND METHODS

2. Materials and Methods

2.1. Cells, culture reagents and tissues

2.1.1. General principles

Cell culture was carried out in a laminar flow hood, which provided a sterile environment. Non-sterile tissue culture reagents were filter-sterilised, using 0.22 µm syringe-driven or vacuum-driven filters (Millipore), and stored in sterile containers at 4°C. All cells were grown in an humidified atmosphere, maintained at 37°C and 5% CO_2. The cells were passaged weekly in T75 flasks (Corning) and all the experiments were performed with sub-confluent cells.

Most of the reagents used for cell culture were provided by the BCI central service. These included 1X PBS, FBS, 10X Trypsin-EDTA, DMEM, Hams-F12 and L-Glutamine (all from PAA Laboratories).

2.1.2. Breast cancer cell lines

The immortalised, non-transformed, epithelial breast cell line MCF10A (Soule *et al*, 1990), the breast adenocarcinoma MCF7 (Soule *et al*, 1973) and the breast ductal carcinoma T47D (Keydar *et al*, 1979) were used predominantly throughout this study. MCF7 and T47D cells were cultured in DMEM supplemented with L-Glutamine and 10% foetal bovine serum (FBS). MCF10A cells required DMEM:Ham's F12 1:1 volume, 10 µg/ml Insulin from bovine pancreas, 500 ng/ml Hydrocortisone, 100 ng/ml cholera enterotoxin and 5% horse serum (all from Sigma). 20 ng/ml of human EGF (Sigma) was added to the MCF10A media following filtration. The MCF10A cell line series composed of MCF10A.neoT, MCF10At1k.c12, MCF10.CA1a, MCF10A.CA1h and MCF10ADCIS.com was also used (Santner *et al*, 2001; Kadota *et al*, 2010). Most grew in the same medium as the original MCF10A cell line except for MCF10A.CA1a and MCF10ADCIS.com, which required DMEM:Ham's F12 1:1 volume and 5% horse serum. Additionally, Cal51, MDA-MB-453, MDA-MB-231, MDA-MB-468, ZR-75-1, β4-1089, BT474, BT20, H3396, SUM159, AU561 and SKBR3 lines were cultured only briefly for genomic DNA purification (Appendix 12).

2.1.3. Fibroblast cell lines

NIH 3T3 fibroblasts were cultured in DMEM and 10% FBS, to use as feeder cells. They were treated with 10 μg/ml of mitomycin C (Sigma) to induce cell cycle arrest. Treated NIH 3T3 cells were seeded in 96 well plates at a concentration of 3,200 cells/well (equivalent to 10,000 cells/cm^2).

2.1.4. Storage and recovery of liquid nitrogen stocks

Mammalian cells were preserved effectively by the presence of a cryoprotectant, dimethylsulphoxide (DMSO), which reduces cellular damage that ice crystals might cause. Cells requiring storage were harvested in log phase growth. Following trypsination, cells were washed in complete medium and resuspended in FBS containing 10% DMSO at 1 to $2x10^6$ cells per ml. One ml of cell suspension was transferred to a cryogenic tube and placed at -80°C in an insulated container for up to two weeks before being transferred to liquid nitrogen (-196°C) for long term storage.

Cryopreserved cells are fragile and require quick thawing and immediate retrieval into complete medium. Vials were therefore placed in a water bath at 37°C to thaw rapidly, and the cell suspension was mixed with 15 ml of complete warm medium and transferred to a T75 culture flask. The medium was changed 24h later, in order to remove any trace of DMSO.

2.1.5. Breast tissue samples

Frozen tissue from ER positive breast tumours was obtained from the Breast Cancer Campaign Tissue Bank (Barts Cancer Institute, BCI), in collaboration with Prof Louise Jones, breast cancer pathologist at BCI, under ethical approval (Ethics REC reference: 10/H0308/49).

2.2. *In vitro* experiments

2.2.1. Proliferation assays

Cell viability and proliferation can be assessed using different approaches, either by detecting proteins only present in proliferating cells (MTS assay and Ki67

staining) or by analysing the different proportion of cells in each phase of the cell cycle based on their DNA content (Flow cytometry with Propidium Iodine staining).

2.2.1.1. MTS assay

Proliferation of cells over a 72 h time period was measured using the CellTiter 96 Aqueous One Solution Cell Proliferation assay (Promega). The assay is a colorimetric method for determining the number of viable cells in culture by incubating then in a solution of a tetrazolium compound, 3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium, and an electron coupling reagent, phenazine methosulphate. Together they react with dehydrogenase enzymes found in metabolically active cells and convert the MTS into formazan. The amount of formazan product, as measured by the amount of absorbance at 490 nm, is directly proportional to the number of living cells in culture.

The cells were seeded on a 96 well plate at a concentration of 2,500 cells/well, in triplicate for each time point (24 h, 48 h and 72 h).

At the end of each time point, the medium was removed and replaced with 100 µl of fresh medium and 20 µl of CellTiter Solution (controls wells did not contain any cells). The plate was incubated at 37°C for 2 h. Absorbance in the different wells was measured at 490 nm on a LT-4000 Microplate reader (Labtech). The wells without cells were used as blanks for normalisation.

2.2.1.2. Ki67 staining

The cells were plated on glass cover slips at a density of 20,000 cells/well in a 24 well plate. The next day, cells were fixed in 4% paraformaldehyde (PFA, Sigma) at room temperature for 10 min and washed three times in PBS for 5 min. The cells were permeabilised in 0.1% Saponin (Sigma) for 10 min, followed by three PBS washes. Non-specific antibody binding was blocked by incubation for 1 h in 5% bovine serum albumin (BSA) in PBS and prior to incubation with anti-Ki67 antibody (FITC Mouse, 1:100 dilution, BD Transduction). The cells were washed several times in PBS with one last wash in distilled water before mounting on a glass slide with mounting media (Prolong™ Gold DAPI antifade reagent, Invitrogen). DAPI

(4',6-diamidino-2-phenylindole), contained in the mounting medium, allowed fluorescent labelling of cell nuclei.

Images were taken on a confocal laser-scanning microscope LSM 510 (Zeiss). Quantification was performed by counting the percentage of Ki67 positive cells per field of view using 40x magnification (10 pictures were used for each cell clone).

2.2.1.3. Cell cycle analysis

After reaching 65% to 80% confluency, the cells were harvested by trypsination, pelleted and resuspended in 1 ml of cold 70% ethanol with vortexing. The cells were fixed at 4°C for 30 min (but could be stored for up to a week) before being processed for staining with propidium iodide (PI, Sigma). After two washes in PBS, the cells were resuspended in 350 µl of staining solution containing 50 µg/ml of PI and 100 µg/ml of RNase (Sigma) diluted in PBS. The tubes were protected from light and incubated at room temperature (RT) for 30 min.

The amount of DNA staining was then assessed by flow cytometry on a FACSCalibur machine (BD Biosciences). Raw data were analysed using FlowJo™ software, using the Watson (Pragmatic) algorithm. Two-way Anova statistical test was used to determine significance (GraphPad Prism, version 5.03).

2.2.2. ERα pathway inhibition

Cells were seeded in 6 well plates at a density of $3x10^5$ cells per well in normal medium. The cells were treated with 1 µM Tamoxifen (Sigma) or with ethanol (vehicle control) for 48 hours and total RNA was purified using an RNeasy Kit (Qiagen) according to manufacturer's recommendations. Complementary DNA was generated from 500 ng of RNA and quantitative real time PCR performed using SYBR Green (Invitrogen) and the StepOnePlus Real-time PCR system (Applied Biosystems).

2.2.3. FGF7 and FGF10 stimulation

Cells were seeded in 6 well plates at a density of $3x10^5$ cells per well in normal medium. The next day, the medium was replaced by starvation medium (DMEM + 0.1% BSA) and the cells were starved overnight. The starved cells were then

stimulated for a period of time ranging from 5 min to 1 hour with several different concentrations of ligand (100, 50, 10, 1 ng/ml of FGF7 or FGF10) and 300 ng/ml of Heparin. At the end of the time course, the cells were lysed in 2X NuPage Sample buffer (Invitrogen) supplemented with 10mM DTT and a western blot was performed using anti phospho-ERK antibody (#9101S, Cell Signalling). Equal loading was verified using Ponceau staining (Sigma) on the membrane prior to antibody incubation.

2.2.4. Selection pressure experiment

MCF7 cells ($2x10^6$ cells) were transfected in triplicate with mRNA encoding the ZFN pairs, along with the MCF7 repair template, as described in section 2.5.2. At passage 1 post-nucleofection, and every third passage thereafter, the cells from each triplicate transfection were divided into 3 different T75 flasks:

1-maintenance

2-genomic DNA extraction

3-Liquid nitrogen stock

The genomic DNA was used for Taqman SNP genotyping assay to determine relative presence of the major and minor allele of rs2981578 SNP over a period of 20 passages.

2.2.5. Single cell dilution and colony picking

Following ZFN transfection, cell screening was performed on a clonal population of cells that were obtained by single cell dilution cloning or colony picking.

2.2.5.1. MCF10A cells

MCF10A cells were isolated using serial dilutions of a cell suspension (Fig. 2.1). 10 to 20 plates were prepared and incubated at 37°C for 7 to 10 days. The single colonies found in some wells were then trypsinised and transferred to a new 96 well plate. Alternatively, a very low concentration of MCF10A cells (4 cells/ml) was used. 100 µl of cell suspension was seeded in each well of a 96 well plate. These cells were co-cultured with inactivated NIH 3T3 feeder fibroblasts, previously treated with mitomycin C. After the MCF10A cell colonies reached 50 to 100 cells,

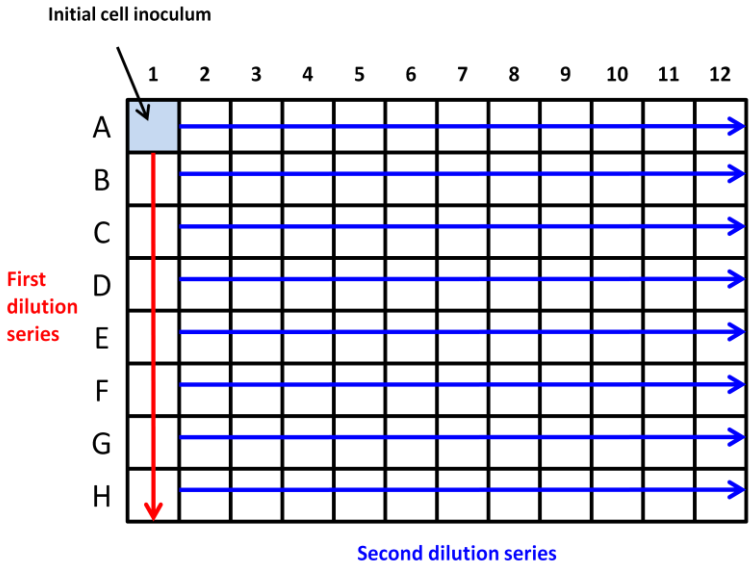


Figure 2.1: Schematic set up of serial dilution cloning in a 96 well plate

Well A1 of a 96 well plate was seeded with $9x10^4$ cells, then serial 1:1 dilutions were performed vertically from A1 to H1 (red arrow). Finally, using a multi-channel pipette, the second serial dilutions were performed horizontally across all the remaining wells in rows 2 to 12 (blue arrows).

they were detached by trypsination and transferred to a new 96 well plate in order to remove any non-cycling feeder cells.

2.2.5.1. MCF7 cells

MCF7 cells were seeded at a concentration of 200 cells/plate in 150 mm diameter culture plates and cultured for a minimum of 7 days. Once the colonies reached approximately 100 cells in size, the medium was removed and the cells washed with sterile PBS. Using a 200 µl tip and/or cloning rings, individual colonies were picked and transferred to a 96 well plate containing fresh medium.

Alternatively, following GFP enrichment (see section 2.5.2), MCF7 cells were diluted to a concentration of 4 cells/ml or single cells were directly FACS-sorted into a 96 well plate. 100 µl of this cell suspension was then seeded into each well of a 96 well plate, with NIH 3T3 feeder cells previously treated with mitomycin C.

2.2.6. Migration assay using Organotypic cultures

Conventional cell culture models involve the culture of cells on two-dimensional (2D) substrates. Breast cancer cells can adapt to this synthetic environment, become flattened and behave in an adherent fashion. Different methods have been developed, however, to study epithelial cell behaviour in a more physiological context, and in the presence of mesenchymal cells. The organotypic culture model adopted consisted of growing epithelial cells on top of a collagen/matrigel substrate embedded with fibroblasts to provide a chemoattractant signal to encourage cell migration and invasion (Fig 2.2). MCF7 cells do not demonstrate invasive behaviour (unpublished data) in this assay, so it was therefore modified to study migration. After 6 days of culture, the cells form a layer covering the whole surface of the gel. This cell layer was subject to punch biopsy wounding, leaving a hole in the middle of the gel. The capacity of the cells to close the wound was then assessed over a period of 14 days.

2.2.6.1. Wound assay

The migratory capacity of modified MCF7 clones was analysed by organotypic culture wound healing assay (Fig. 2.2). This was modified from previously published protocols (Nystrom *et al*, 2005; Chioni *et al*, 2010).

3.48 mg/ml collagen type I (Millipore) and Matrigel (BD) were mixed in a ratio of 70:30 (80% of final gel volume), 10× Hank's buffer (10% final gel volume) was added to the mix, and pH was adjusted to 7.4 with 2 M NaOH. Human foreskin fibroblasts (HFF2) were resuspended in FBS (10% final gel volume) at a concentration of 5×10^5/ml and added to the mix. The final mixture was applied to a 24-well plate (1 ml/well) and incubated at 37°C in 8% CO2 for 4 h, after which the gels were equilibrated by immersion in medium for 16 h, whereupon the medium was replaced by 500 µl culture medium containing $1x10^6$ cells of MCF7 derived clones. 250 µl collagen mix (7 vol collagen type I, 1 vol each of 10× Hank's buffer, FBS, and culture medium neutralised with 2 M NaOH) was added dropwise onto 400 mm^2 Nylon membranes (100-µm pore; Tetko). Membranes were incubated at 37°C for 30 min and then fixed for 1 h at 4°C with 1% glutaraldehyde (Sigma-Aldrich)/PBS. After fixation, the membranes were washed 4 times for 5 min in PBS and incubated overnight in culture medium at 4°C. The coated membranes were placed on 25 mm^2 sterile stainless steel grids in 6-well plates. Gels were lifted from the 24-well plate and laid on top of the coated membranes. An appropriate amount of culture medium was added to each well until it reached the lower part of the gel, so that the cultures were maintained at the air-liquid interface. Medium was changed every 2 days. After 6 days of culture, the thin layer of cancer cells proliferating on the top of the organotypic gel was wounded using a circular Biopsy punch (Stiefel, 5 mm) and all the cells of the wound were removed. The organotypic cultures were maintained for different periods of time (0, 9 and 14 days post-wounding). The gels were fixed in 10% neutral buffered Formalin (CellPath) for 16 h at 4°C. After fixation, gels were washed thoroughly in PBS, bisected, and dehydrated through a graded ethanol series before wax embedding.

The sizes of the wounds (in triplicate) at the end of the experiment were assessed using light microscopy and compared to the initial wound size (day 0).

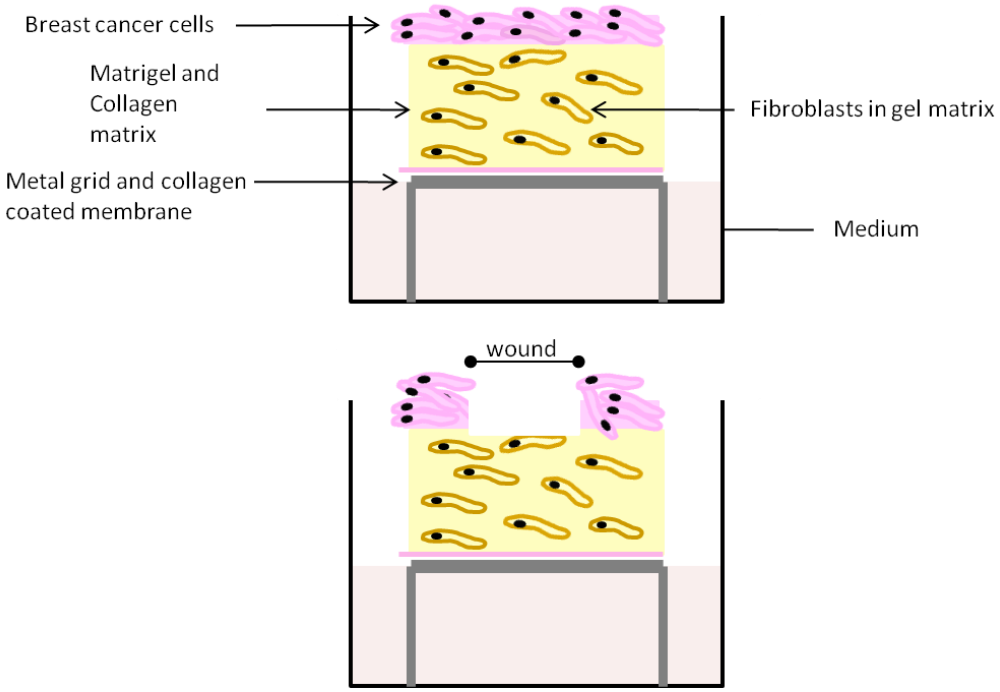


Figure 2.2: Wound assay: Organotypic culture to study cell migration

The cancer cells were seeded (pink) on top of a matrigel and collagen matrix (yellow) containing stromal cells (fibroblasts). The gels were raised to the air-liquid interface, on a metal grid (grey) and fed with normal growth medium from underneath. To study cell migration, the cell layer was wounded and the closure of the wound monitored, at several time points.

2.3. DNA

2.3.1. Genomic DNA extraction

Nucleic acids such as DNA have the capacity to bind to solid substrates (silica membrane) at specific pH and high salt concentration. This property forms the basis of most of the current methods to extract and purify DNA, in which the cell lysates are passed through a colomn with a silica membrane, washed with ethanol, and eluted in a low salt solution. Another method routinely used to purify DNA is based on the interaction of DNA with a mixed solution of phenol and chloroform. Upon addition of equal amounts of phenol and chloroform to a cell lysate, a biphasic mixture is formed and, after centrifugation, results in an upper aqueous phase (containing the DNA) and a lower organic phase, with proteins present at the interphase. The nucleic acids contained in the upper phase are then precipitated using ethanol. This method achieves a high DNA purity but is more time consuming than the silica column based technique.

2.3.1.1. Cell lines

Total DNA from breast cancer cells was extracted using the GenElute^{TM} mammalian genomic DNA miniprep kit (Sigma), according to the manufacturer's instructions. Cultured cells ($2x10^6$ to $4x10^6$ cells) were detached with Trypsin, and pelleted by centrifugation at 300x *g* for 5 min. The pellet was resuspended in 200 µl resuspension solution and digested in 180 µl Lysis buffer C supplemented with 20 µl proteinase K (20 mg/ml) and incubated at 70°C for 10 min. 200 µl 100% ethanol was added to the lysate and transferred to the pre-equilibrated binding column. Wide bore pipette tips were used to reduce DNA shearing. The columns were centrifuged at 6,500x *g* for 1 min, washed twice with wash solution containing 100% ethanol and eluted in 200 µl of elution solution.

2.3.1.2. Tissues

Total DNA from breast tissues was extracted using the GenElute^{TM} mammalian genomic DNA miniprep kit (Sigma Aldrich) as above but using Lysis buffer T instead and digested for up to 4 hours at 55°C.

The gDNA concentration was measured using a Nano Drop spectrophotometer (ND-1000).

2.3.1. DNA purification

2.3.1.1. QIAquick kit

PCR products or ChIP samples were purified using a QIAquick PCR purification kit (Qiagen), according to the manufacturer's instructions. Five volumes of PB buffer were added to one volume of PCR product. The pH was adjusted using 3M sodium acetate in order to keep the pH indicator yellow and the mix was transferred to a QIAquick spin column. The DNA was bound to the column by centrifugation at 16,000 x *g* for 1 min followed by a wash with 750 µl buffer PE and a centrifugation at 16,000 x *g* for 1 min. DNA was eluted by applying 50 µl EB buffer (10 mM Tris-Cl, pH 8.5) to the column and centrifugation for 1 min.

2.3.1.2. Phenol/chloroform method

DNA was mixed with phenol/chloroform isoamyl alcohol (25:24:1) at a 1:1 v/v ratio. The solution was vortexed thoroughly and spun at 16,000x *g* for 3 min. The upper phase was transferred into a fresh tube, with the addition of an equal volume of 3M sodium acetate and 2.5 volumes of 100% Ethanol. The mixture was incubated 1h on dry ice or overnight at -20°C and the DNA was pelleted by centrifugation at 16,000x *g*, 4°C for 20 min. The DNA pellet was washed once with 70% ethanol, air-dried and resuspended in an appropriate volume of sterile distilled water.

2.3.1.3. Gel electrophoresis

DNA fragments were separated on 0.8% w/v or 1.5% w/v agarose gels made with UltraPureTM agarose (Invitrogen) in 1X TAE with 1X Gel Red (10,000X, Biotum). Loading buffer was added to samples prior to loading (6X, Fermentas). Bands were visualised under UV light and photographed using an AutoChemi System (UVP BioImaging System).

2.3.1.4. Gel extraction

DNA was extracted from agarose gels using the QIAquick Gel Extraction Kit (Qiagen) according to manufacturer's instructions. The desired band was excised using a scalpel and dissolved in 3 volumes of buffer QG for 10 min at 50°C. 1 gel volume of isopropanol was added and the mix was transferred to a QIAquick spin column. The DNA was bound to the column by centrifugation at 16,000 x *g* for 1 min followed by a wash with 750 µl buffer PE and a centrifugation at 16,000 x *g* for 1 min. DNA was eluted by applying 30 µl of elution buffer to the column and centrifugation for 1 min.

2.3.2. Cloning

All restriction enzymes and enzyme buffers used in this study, unless stated otherwise, were obtained from New England Biolabs.

2.3.2.1. DNA repair templates

a. Double-stranded DNA repair template

The DNA sequence that was used as a repair template after the ZFN-mediated cut required a 1 kb homology arm each side of the rs2981578 SNP. The 2kb fragment was obtained by PCR using genomic DNA from MCF7 and MCF10A cell lines with hFGFR2_11496-11515_for and hFGFR2_13647-13628_rev primers (Primer table, section 2.8). The following PCR conditions were used:

Megamix (Microzone)	47 µl
Primers mix (100mM)	0.5µl
gDNA	2.5 µl

94°C	2 min	
94°C	30 s	30 cycles
60°C	30 s	
72°C	2 min	
72°C	7min	

The purified PCR product was cloned in pJET1.2/blunt vector (Fermentas) as follows:

2X reaction buffer	10 µl
PCR product	2 µl
Water	up to 17 µl
DNA blunting enzyme	1 µl

The mixture was incubated at 70°C for 5 min and chilled on ice. The following was added to the blunting reaction:

pJET1.2/blunt cloning vector (50 ng/µl)	1 µl
T4 DNA ligase (5 units/µl)	1 µl

The ligation reaction was incubated at RT for 5 min and used for bacterial transformation of *E. coli*, DH5α strain (Bioline, α-select, Gold Efficiency).

b. Single stranded DNA repair template

A second type of repair template, containing the risk allele (bold and underlined) was used to edit the MCF7-derived clones using the FGFR2 ZFN pair. The 137 base repair template (below) was synthesized by Integrated DNA technologies (IDT) and transfected (2 µg) into cells, similarly to the double-stranded DNA repair template (section 2.5.2).

5'-CAGCCCTTCTGAGATCTAAAGCTTCCCTCTGAATGCTGCTTTGGAGGATTGTGAGAGG TAGTGACTCTTCAAAGTTTGTTTGTTTTCTTGAAGCTTTTACCTCTATGCAAATATGC**G**GTT TGGAGCAGGGAAGAAA-3'

2.3.3. Bacterial transformation

DNA (5 µl ligation mixture or 1-10 ng plasmid DNA) was added to 50 µl chemically competent bacteria (Bioline, α-select, Gold Efficiency), incubated on ice for 15 min, heat shocked for 30 s at 42°C and placed on ice for 2 min. The bacteria were resuspended in 500 µl antibiotic-free warm Luria Broth (LB) and incubated at 37°C with shaking at 225 rpm for 1h. 100 µL to 200 µl cells were plated onto LB agar plates containing the appropriate selection antibiotic (100 µg/ml) and incubated overnight at 37°C.

2.3.4. Preparation of plasmid DNA

Small scale plasmid DNA preparations were carried out using a QIAprep Spin Miniprep kit (Qiagen) according to the manufacturer's instructions. A 2 ml overnight culture was pelleted by centrifugation and the pellet was resuspended in 250 µl cell resuspension solution (P1). 250 µl cell lysis solution (P2) was added and mixed by inversion; 350 µl neutralisation solution (N3) was added and mixed by inversion. The mix was centrifuged at 16,000 x *g* for 10 min and the clear lysate was transferred into a spin column and spun at 16,000 x *g* for 1 min. The column was washed by adding 750 µl wash solution (PE), spun at 16,000 x *g* for 1 min, followed by a 2 min spin at maximum speed to dry the column. DNA was eluted by adding 50 µl nuclease-free water or EB buffer to the column and stored at -20°C.

Large scale plasmid DNA preparations were carried out using the QIAfilter plasmid purification kit (Qiagen) according to manufacturer's instructions. A 200 ml overnight culture was pelleted by centrifugation (15 min at 6,000x *g*) and the pellet was resuspended in 10 ml buffer P1. 10 ml buffer P2 was added, mixed by inversion and incubated at RT for 5 min. 10 ml chilled buffer P3 was added and mixed by inversion before the lysate was poured into a QIAfilter cartridge and incubated for 10 min at RT. The clear lysate was transferred into a QIAGEN-tip (pre-equilibriated with 10 ml buffer QBT) and entered the resin by gravity flow. The QIAGEN-tip was washed twice with 30 ml buffer QC. DNA was eluted with 15 ml buffer QF and precipitated with 10.5 ml isopropanol. The solution was centrifuged 30 min at 15,000x *g* and the DNA pellet washed with 5 ml 70% ethanol and centrifuged 10 min at 15,000x *g*. DNA pellet was air-dried under the hood and eluted in 1 ml sterile distilled water and stored at -20°C. DNA concentrations were measured on a Nanodrop spectrophotometer.

2.3.1. Surveyor assay

The Surveyor assay provides a way of visualising mutations in DNA after PCR, endonuclease digestion and resolution on an agarose gel. *Cel-I* nuclease, a mismatch-specific endonuclease derived from celery, can recognise and cleave all types of mismatches arising from the presence of single nucleotide polymorphisms

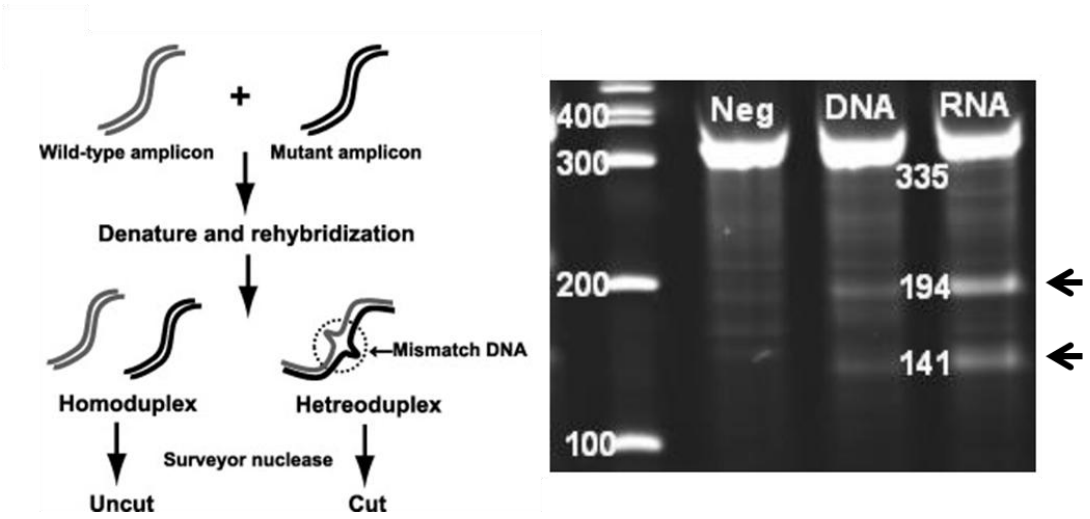


Figure 2.3: Surveyor nuclease assay for detection of ZFN activity

PCR on ZFN-edited cells generates a mixture of wild-type and mutated PCR products (left). Melting and re-annealing of the PCR strands creates a mixture of homoduplex and heteroduplex DNA where mismatches are present (circled) and are substrates for the Surveyor nuclease (*Cel-I*). Electrophoresis gel: lane 1-Cells transfected with a GFP plasmid (Neg); lane 2- cells transfected with plasmids encoding ZFNs (DNA); lane 3- cells transfected with ZFNs mRNAs (RNA). Arrows point to bands of the expected sizes based on the size of the PCR product and the site of ZFN target site cleavage (gel from *FGFR2* ZFN certificate of Analysis, Appendix 5).

(SNPs) or from small insertions or deletions (Fig. 2.3). Purified gDNA from cells transfected with zinc finger nuclease (ZFN) mRNA or pmaxGFP vector (Lonza) was amplified using ZFN specific primers (25µM) (Primer table, section 2.8) with the following reaction conditions:

Initial denaturation	95°C	3 min	
Denaturation	95°C	30 s	30 cycles
Annealing	57°C	30 s	
Extension	72°C	30 s	
Final extension	72°C	7 min	

After PCR, the product was denatured and re-annealed to create mismatch duplexes using the following thermal cycler conditions:

95°C	10 min
85°C	Cool at 2°C/second
25°C	Cool at 0.1°C/second
4°C until next step	

The *Cel-I* endonuclease from the Surveyor™ mutation detection kit (Transgenomic) was then used to cleave DNA at any mismatch bubble. 1 µl Surveyor endonuclease and 1 µl Surveyor enhancer were added to 10 µl re-annealed PCR products and the sample was then incubated at 42°C for 45 minutes. The product was resolved on a 2.5% high resolution agarose (Sigma) gel stained with 1X Gel Red (Biotum).

2.3.2. Genotyping

Sanger sequencing, a method which incorporates fluorescent dideoxynucleotides (ddNTPs) into newly synthetised DNA during *in vitro* DNA replication, was used to assess the SNP status of different breast cell lines.

2.3.2.1. Sequencing of plasmid DNA

For assessing the SNP status of rs2981578 in different breast cell lines, a 500 bp insert generated by PCR using SNP_For and SNP_Rev primers (Primer table, section

2.8) was cloned in pJET1.2/blunt vector using the CloneJET PCR cloning kit, transformed into chemically competent *E.coli* (DH5α strain) and at least 6 colonies were sequenced (Genome Centre, BCI) using the same primers, following preparation of plasmid DNA. Sequencing files were analysed using BioEdit Sequence Alignment editor (version 7.0.5.3) and CLC Sequence viewer 6 software.

2.3.2.2. Cycle sequencing of PCR product

Cycle sequencing was performed to genotype several *FGFR2* SNPs. The region containing the SNPs was first amplified by PCR using a proof reading DNA polymerase, Hot Start Taq DNA polymerase (Qiagen), in the following reaction mixture:

PCR buffer 10X w or w/o Coral load (Qiagen)	2 ul
dNTPs (2mM) dilution 1:5	2 ul
Primer 1 (10 µM)	1 ul
Primer 2 (10 µM)	1 ul
Hot start Taq polymerase	0.1ul
DNA template	1 ul
Water	13 ul

The PCR programme conditions were as followed:

94°C	5 min	
94°C	30 s	35 cycles
58°C	30 s	
72°C	30s	
72°C	10 min	

The samples were cleaned using Exonuclease 1 (Exo) (20,000 units/ml, New England Biolabs) and Shrimp Alkaline phosphatase (SAP) (1 unit/µl, USB).

Exo	1 µl
SAP	20 µl
Water	179 µl

4 µl of the above mix was added to each 5 µl PCR sample, and incubated for 15 min at 37°C, followed by 15 min at 80°C in order to inactivate the enzymes.
BigDye (Applied Biosystems) sequencing was performed using one of the original primers used for PCR or an internal primer, when specified.

BigDye v3.1	0.25 µl
5X Buffer	1.875 µl
Water	6.375 µl
Primer (3.2 µM)	0.5 µl
Cleaned PCR sample	1 µl

The following conditions were used:

96°C	1 min	
96°C	30 s	25 cycles
50°C	15 s	
60°C	4 min	

The post-PCR reaction mixture was sent to the Barts Genome Centre for capillary electrophoresis and the resulting sequencing trace was returned to us for analysis. Sequencing files were analysed using BioEdit Sequence Alignment editor (version 7.0.5.3) and CLC Sequence viewer 6 software.

2.3.2.1. Allelic discrimination using Taqman assays

SNP genotyping using Taqman probes allow the direct genotyping of biallelic SNPs directly from gDNA without the need for cloning the SNP locus prior to the genotyping step. It can discriminate homozygous from heterozygous samples by the use of two fluorescent probes. The genotypes of rs2981578, rs1047100 and rs755793 (Primer table, section 2.8) were determined by Taqman SNP genotyping assay (Applied Biosystems) in a panel of breast cancer cell lines and breast tissue samples (Fig. 3.1A and Table 5.1). Each sample was used in duplicate in the following reaction mix:

Taqman SNP genotyping assay 40X	0.25 µl
Taqman genotyping master mix 2X	5 µl
Water	3.75 µl

AD files were used to visualise the genotyping results (Genotyper software, version 1.0.1, Applied Biosystems), whereas specific allele amplification data could be read with RQ files, using SDS software, version 2.3 (Applied Biosystems).

2.3.3. Site-directed mutagenesis (SDM)

Site-directed mutagenesis is a technique used to make point mutations on a circular double stranded plasmid DNA, witch ultimately allows the substitution, deletion or insertion of single or multiple amino acids in a recombinant protein (Fig. 2.4, step 1). It requires of pair of complementary primers that contain the mutation to be introduced in the wild-type plasmid sequence (step 2). After a PCR amplification of the mutated plasmid, the product is digested by the endonuclease *Dpn I*, that can recognise and digest the methylated, wild-type DNA and leave the newly synthesized DNA untouched (step 3). The final product is then cloned into competent bacteria to be recircularised and amplified (step 4).

Site directed mutagenesis was performed using an adaptation of the QuikChange protocol (Fig. 2.4) (Stratagene). KOD DNA polymerase (Novagen) was used to amplify 50 ng DNA template (2kb of *FGFR2* intron 2 containing the SNP and cloned in pJET1.2/blunt vector) using mutagenesis primers created using the QuikChange™ Primer Design Programme (Agilent Technologies).

DMSO	1 µl
10X Buffer #2	5 µl
$MgCl_2$ (25 mM)	3 µl
dNTPs (2 mM each)	5 µl
Forward primer (5 µM)	4 µl
Reverse primer (5 µM)	4 µl
Template DNA	50 ng
KOD DNA polymerase	0.4 µl
Water	up to 50 µl

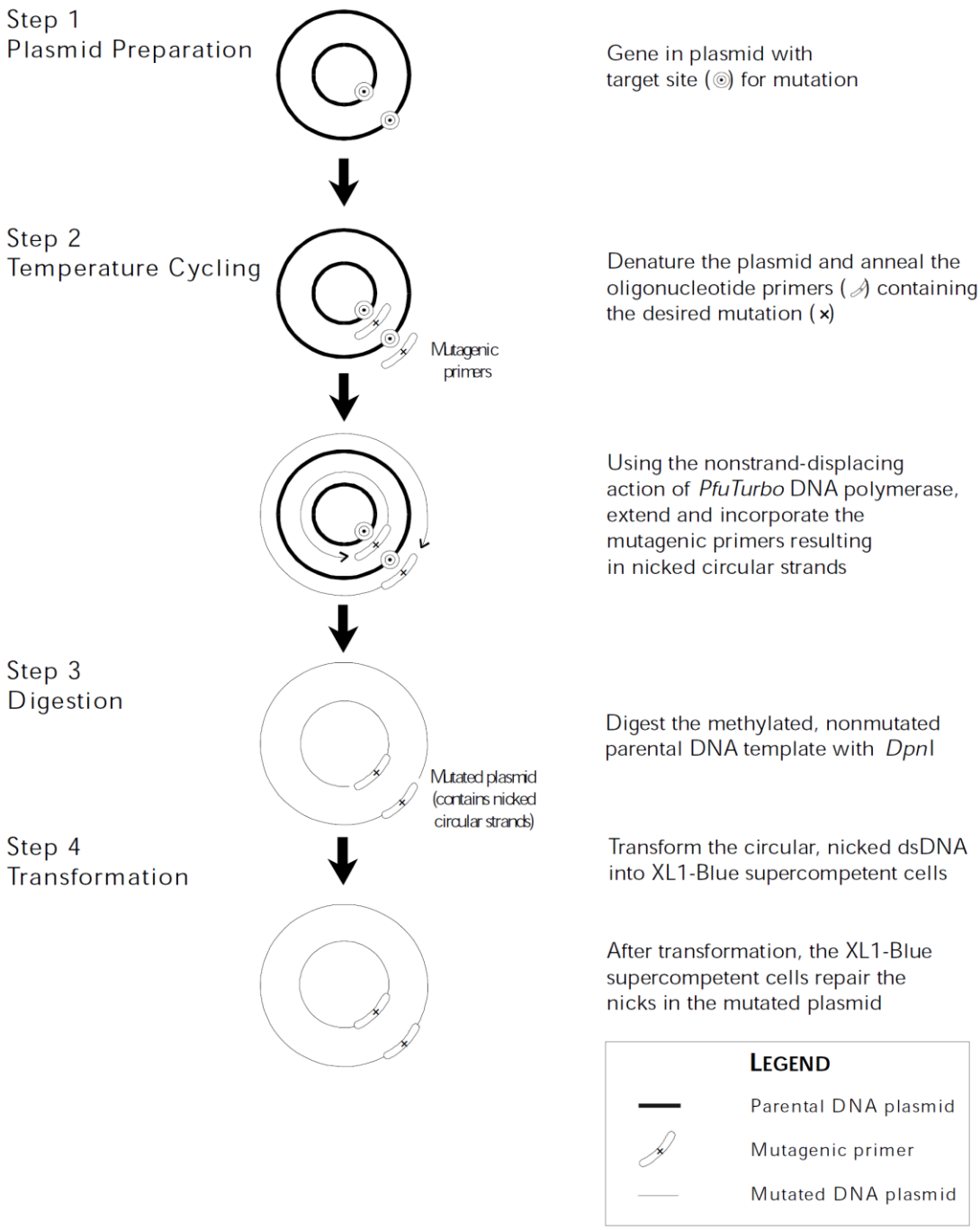


Figure 2.4 : Overview of the different steps involved in Site-directed Mutagenesis (SDM)

Site-directed mutagenesis is a multistep molecular biology technique used to specifically mutate a DNA plasmid using mutagenesis primers and bacterial cloning. Source : Quick Change Site-directed mutagenesis kit, Stratagene.

94°C	15s	
60°C	30 s	18 cycles
72°C	1 m 40 s	

The amplification was followed by 2 to 3 hours of *DpnI* digestion to remove any methylated parental DNA, and the remaining DNA was transformed in competent DH5α *E. coli* cells, prior to preparation of plasmid DNA from overnight cultures.

2.4. RNA

2.4.1. RNA isolation

The same principles as gDNA extraction and purification apply for RNA.

2.4.1.1. Cell lines

Total RNA was isolated using an RNeasy Kit (Invitrogen) according to manufacturer's recommendations for cells cultured as monolayer. RTL buffer (350 µl/well of a 6 well plate) was added to the washed cell monolayers and the cells detached from the flask using a cell scraper. The cell lysate was homogenised by several passes through a 0.8 mm diameter sterile needle and 350 µl of 70% ethanol was added before transfering the lysate to the columns. The column was centrifuged for 30 s at 8,000x *g*, 350 µl RW1 buffer was added before being spun again. DNase treatment (Qiagen) was performed by incubating the membrane with a mixture of 50 µl DNase and 350 µl RDD buffer for 15 min at RT. The column was washed twice with RPE buffer and the RNA was eluted in 50 µl RNase-free distilled water. RNA concentration was determined by the absorbance measurements at 260/280 nm using a Nanodrop spectrophotometer.

2.4.1.2. Breast tissues

Total RNA from breast cancer tissues, previously snap frozen in liquid nitrogen, was purified with Trifast reagent (PeqLab). The tissue sample (80 to 100 mg) was immersed in 1.5 ml of Trifast and homogenised using a tissue homogeniser (Ultra-Turrax), before being incubated for 5 min at RT. 300 µl chloroform (Fisher Scientific) was added and the tube shaken vigorously, before 3 min incubation at

RT. The different phases were separated by 5 min centrifugation at 12,000x *g*. The upper aqueous phase containing the RNA was transferred to a fresh tube and RNA was precipitated by addition of 750 µl isopropanol (Fisher Scientific). The samples were kept on ice for 15 min and centrifuged at 4°C, 12,000x *g* for 10 min. The RNA pellets were washed twice with 75% ethanol, mixed by vortexing and spun at 12,000x g for 10 min (4°C). The excess liquid was removed and the pellets were air-dried and resuspended in 200 µl RNase-free water. The samples were then stored at -80°C.

2.4.2. cDNA synthesis

Complementary DNA was generated by reverse transcription from RNA isolated from tissues or cell lines. Random hexamer primers were used to prime the reverse transcription reaction and ensure the efficient amplification of all mRNAs in the samples and avoid the bias toward the 3' end of message, seen when using oligo dT primers (Bookout and Mangelsdorf, 2003). 0.5 µl of random hexamers (50 ng/µl) were mixed with 400 ng (or up to 1 µg) total RNA, plus 1 µl dNTP Mix (10mM each) in a total volume of 12 µl (topped up with distilled water). The mixture was heated to 65°C for 5 min and then quickly chilled on ice. The contents of the tube were collected by brief centrifugation, then 4 µl of 5X First-Strand buffer and 2 µl DTT (0.1M) were added.

After incubation at RT for 2 min, 1 µl Superscript^{TM} II Reverse transcriptase enzyme (200 units, Invitrogen) was added. The final mixture was incubated at 25°C for 10 min to allow primer binding, then at 42°C for 50 min during which period the reverse transcription occurred. A final 15 min at 70°C was required to inactivate the enzyme. cDNA samples were used immediately for RT-PCR or Taqman assay whenever possible, or stored at -20°C.

2.4.3. Real time polymerase chain reaction

All real time PCR reactions were carried out in triplicate in a 20 µl reaction volume containing 0.06 µl of each primer (100 µM stock), 1 µl cDNA, 10 µl SYBR green (Qiagen) and 8.88 µl RNase-free distilled water.

Serial dilutions (1, 0.75, 0.5, 0.25, 0.1, 0.01, 0.001) of one of the control samples were used to establish a standard curve for each of the new primer sets used in the real time PCR reaction. GAPDH and/or HPRT primers were used as internal controls.

PCR amplification was performed using a StepOne® (Applied Biosystems) Real-Time PCR system under the following thermocycling profile:

Reaction conditions:

95°C	15 min	
95°C	30 s	40 cycles
60°C	30 s	
72°C	30 s	
Melting curve		
95°C	60 s	
60°C	30 s	
95°C	60 s	

The results were analysed using the Comparative $^{\Delta\Delta}$CT method and presented as the mean of three independent experiments (Livak and Schmittgen, 2001).

2.4.4. miRNA isolation and amplification by q-RT-PCR

Micro RNAs (miRNA) are small, single stranded, non-coding RNA molecules that play a role in transcriptional and post-transcriptional regulation of gene expression. Their main function is the downregulation of gene expression, either by repressing mRNA translation or by targeting the mRNA for degradation. They originate from pre and pri-mRNA which are processed into miRNA by Drosha and the miRNA-RISC Complex (Chen and Rajewsky, 2007).

A mirVana isolation kit (Applied Biosystems) was used to amplify the total miRNA population expressed in MCF10A and MCF7 cells. $2x10^6$ cells were trypsinised, washed in PBS and pelleted. The pellet was lysed on ice by addition of 600 µl lysis/binding solution and vortexed to mix. 1:10 v/v miRNA homogenate additive was added and incubated on ice for 10 min. 600 µl Acid-Phenol/Chloroform

solution was then added and the solution vortexed for 1 min. The aqueous and organic phases were separated by 5 min centrifugation at 10,000x *g*. The upper (aqueous) phase was transferred into a fresh tube and 1.25 volumes of 100% ethanol at RT were added. The solution was pipetted into a filter cartridge, spun for 15 s and washed with 700 µl wash solution 1. The wash was repeated with 500 µl wash solution 2/3 and the RNA was eluted by applying 100 µl pre-heated (90°C) elution buffer. The total miRNA solution was kept at -20°C. The miRNA was reverse transcribed using specific primers for miR-221 and miR-19a (Taqman microRNA assays, Applied Biosystems) and the Taqman MicroRNA Reverse transcription kit (Applied Biosystems) according to the following reaction:

dNTPs (100 mM)	0.15 µl
Reverse transcriptase enzyme (50 units/µl)	1 µl
10X Reverse transcription buffer	1.5 µl
RNase inhibitor (20 units/µl)	0.19 µl
RNase-free water	4.16 µl

The master mix was transferred to PCR tubes, 5 µl total miRNA (150 ng/µl) and 3 µl 5X stem loop miR primers (miR-221 or miR-19a) were added. The reaction was incubated on ice for 5 min and transferred to a thermocycler:

16°C	30 min
42°C	30 min
85°C	5 min

The reverse transcription reaction was used in quadruple for RT-PCR:

2X Taqman master mix	10 µl
Taqman primer+ probes	1 µl
RNase-free water	8 µl
miR cDNA	1 µl

The following programme was used:

95°C	10 min	
95°C	30 s	40 cycles
60°C	1 min	
4°C	hold	

2.4.5. Custom-made *FGFR2* zinc finger nucleases

The CompoZr™ custom made *FGFR2* ZFNs were purchased from Sigma-Aldrich. Vials of ZFNs mRNA sufficient for 10 transfections were provided. Additional mRNA was generated from ZFN plasmids (Sigma). Briefly, the two plasmids were transformed in *E. coli* (using Kanamycin selection) and purified. 20 µg of each ZFN plasmid was digested with 10 µl of *XbaI* to linearise the plasmid template (Appendix 4). The linearised DNA was purified by phenol chloroform extraction and used for mRNA run-off transcription using MessageMax T7 mRNA transcription kit (Epicentre).

2.4.5.1. mRNA synthesis

Messenger RNA can be synthesised *in vitro* by run-off transcription from an expression vector containing a T7 promoter and nascent molecules can be stabilised by addition of a 5' cap and 3' poly A tail (Appendix 4).

The following reagents were combined in a 1.5 ml eppendorf tube for each ZFN plasmid:

RNase-free water	5 µl
Prepared plasmid template.	1 µl
10X transcription buffer	2 µl
Cap/NTP premix	8 µl
DTT (100 mM)	2 µl
MessageMax T7 Enzyme solution	2 µl

The mixture was incubated at 37°C for 30 min, before the addition of 1 µl of DNase and another incubation at 37°C for 15 min.

2.4.5.1. PolyA tailing

An mRNA PolyA tailing reaction was set up using an A-Plus Poly(A) Polymerase tailing kit (Epicentre) as follows:

RNase-free water	55.5 µl
10X A-plus reaction buffer	10 µl
ATP (10 mM)	10 µl
ScriptGard RNase inhibitor	2.5 µl
In vitro transcription reaction (freshly prepared)	20 µl
A-Plus Poly(A) polymerase	2 µl

The mixture was incubated at 37°C for 30 min.

The final mRNA transcripts were purified with MEGAClear kit (Ambion) according to the manufacturer's instructions. To the poly(A) tailing reaction, 350 µl binding solution concentrate and 100 µl 100% ethanol was added. The solutions were then transferred to filter cartridges and centrifuged for 1 min at 15,000x *g*. The cartridges were washed with 500 µl wash solution and spun at 15,000x *g* for 1 min. The RNA was eluted twice in two separate tubes by adding 50 µl elution solution and incubating the cartridge at 65°C for 10 min. The two eluates were pooled together under the hood, on ice, in aliquots of 2 µg each, ready to be used in transfection. The vials were stored at -80°C.

2.4.6. RNA quality control for Breast tissue samples

The quality of RNA purified from snap frozen breast tissue samples was assessed using an RNA 6000 Nano Kit (Agilent Technologies). The assay was performed by the Barts Genome Centre. The results are displayed as RNA concentrations, ribosomal ratio between the intensity of the signal from ribosomal RNA subunits 18S and 28S and the RNA Integrity Number (RIN). The RIN software algorithm allows for the classification of eukaryotic total RNA, based on a numbering system from 1 to 10, 10 being the most intact. Samples below 2 were excluded from analysis.

2.5. DNA and RNA transfection

Nucleic acids can be transfected into mammalian cells using different techniques which are based on two distinct principles. The first is lipid-based transfection, in which the nucleic acid is enclosed in a micelle formed by lipids, that can fuse with the lipid bilayer of eukaryotic cells and therefore incorporate the nucleic acid in the cytoplasm. The second technique, electroporation (nucleofection), is more toxic to the cells but allows higher transfection efficiency. The nucleic acids are present in solution with the cells and get incorporated in the cytoplasm by the action of an electric current, applied for a very brief moment.

2.5.1. Lipid based transfection

a. Lipofectamine

Cells were seeded at $5x10^5$ or $1x10^6$ cells per well in a 6 well plate and cultured for 24 hours to reach 80% to 90% confluency. Lipid-based transfection was performed using 3 µl of Lipofectamine 2000 (Invitrogen) diluted in 250 µl Opti-MEM + Glutamax (Invitrogen). Transfection was performed as per manufacturer's instructions. Briefly, MCF10A and MCF7 cells were transfected with 2 µg either ZFN mRNAs or ZFN plasmid DNAs along with 2 µg repair template (major or minor allele as appropriate). Cells transfected with 2 µg pmaxGFP (Lonza) served as a transfection control. The cells were incubated at 37°C and, 4 hours later, 2 ml complete medium was added. Transfected cells were cultured for 7 days to allow the degradation of the ZFN mRNA and repair template and then used for the establishment of clonal populations of cells.

b. Interferin

Cells were seeded at $3x10^5$ cells per well for MCF10A cells and $1.5x10^5$ cells per well for MCF7 cells in a 6 well plate and cultured for 24 hours to reach 50% confluency. Lipid-based transfection was performed using 4 µl of Interferin (Polyplus) diluted in 100 µl Opti-MEM. Transfection was performed as per manufacturer's instructions. Briefly, MCF10A and MCF7 cells were transfected with 20 µM of siRNA against Oct1 and Runx2. The cells were incubated at 37°C in

1ml complete medium. Total mRNAs and proteins were isolated 48h post-transfection.

2.5.2. Nucleofection

a. ZFN

The ZFN pairs were transfected into breast cancer cell lines using the Amaxa System (Lonza). Nucleofection was performed using the Cell line Nucleofector™ kit (Lonza) according to the following conditions:

Cell line	Kit	Programme
MCF10A	L	T-020
MCF7	V	P-020
T47D	V	X-005

$2x10^6$ cells were pelleted and resuspended in 100 µl transfection solution (with complement), 2 µg plasmid DNA and 2 µg ZFN mRNA. The cell suspension was then transferred into an electroporating cuvette and placed in the nuceofection machine. Immediately after the chosen programme was executed, 500 µl warm complete medium was added to the cuvette and the cell suspension was transferred to a 10 cm culture dish, with 10 ml warm complete medium. The medium was changed 24 h post-Nucleofection, and every other day thereafter.
If GFP enrichment was required, this was performed 48 hours post transfection, which constitutes the peak expression window for the pmaxGFP construct (Lonza). The cells were sorted by Dr Guglielmo Rosignoli or Mr William Day (FACS facility manager and assistant) using the ARIA II cell sorter (Becton Dickinson). Different cell populations were gated to exclude debris, doublets and GFP negative cells. The top 10% to 50% of GFP expressing cells were then collected in a separate culture dish (Appendix 13).

b. Antagomirs

Antagomirs are small synthetic RNA molecules capable of inhibiting the action of miRNA. Their inhibition is very specific toward a single miRNA as the antagomirs are synthesized to be complementary to them. They harbour mutations that reduce their degradation by the cell machinery. The antagomirs (antagomir-

miR221, antagomir-miR222) were transfected into the MCF10A cell line using the Amaxa System (Lonza). Nucleofection was performed using the Cell line Nucleofector™ kit (Lonza) according to the following conditions:

Cell line	Kit	Programme
MCF10A	L	T-020

$2x10^6$ cells were pelleted and resuspended in 100 µl transfection solution (with complement) and 100nM antagormir and electroporated using the Amaxa system. The medium was changed 24 h post-Nucleofection, and every other day thereafter.

2.6. Chromatin Immunoprecipitation (ChIP)

Chromatin immunoprecipitation (ChIP) can be used to determine whether, *in vivo*, a given protein binds to specific regions of DNA within the genome. The cells are fixed using paraformaldehyde, which locks any proteins bound to the DNA in place. The genomic DNA is then purified and sheared into small fragments (200-600 bp) which are used for immunoprecipitation with an antibody against the protein of interest. After only one type of protein/DNA complex has been retrieved, the cross-linking is reversed and the resulting DNA can be analysed by quantitative PCR.

Runx2 ChIP

ChIP for RUNX2 was carried out on target cells using the Magna ChIP™ A/G kit (Millipore) according to manufacturer's instruction. Cells were plated in a 150 mm culture dish and medium was changed every other day until they reached 90% confluence ($1x10^7$ cells in total). The cells were then cross-linked at RT in 1% formaldehyde (Sigma) for 10 min. The reaction was quenched by adding 2 ml 10X glycine at RT for 5 min. The cells were washed twice in cold PBS containing 1X protease Inhibitor cocktail II, scraped and collected by centrifugation at 800x *g* for 5 min (all centrifugation steps were carried out at 4°C). The pellet was lysed on ice for 15 min in Cell lysis buffer containing 1X protease Inhibitor cocktail II. The debris was pelleted by centrifugation and the supernatant was resuspended in 500 µl

Nuclear Lysis buffer. The DNA contained in the lysate was sheared using a Bioruptor (Diagenode) according to the programme [30 s shearing, 30 s stop] repeated 6 or 10 times, as determined by optimisation of the protocol. At this stage the samples were stored at -80°C to be used within 1 month.

50 µl each sample was diluted in 450 µl Dilution buffer, containing protease Inhibitor cocktail II (5 µl was kept as 'input DNA') and 20 µl protein A/G magnetic beads, as well as the immunoprecipitating antibody. 4 µg anti-Runx2 antibody (sc-12488X, Santa Cruz), 1 µg anti-RNA polymerase antibody (05-623, Upstate) and 1 µg IgG control (556648, BD Pharmingen) were used and incubated overnight, at 4°C with rotation. The magnetic beads were then washed in a series of cold buffers and the cross-linking was reversed by addition of ChIP elution buffer containing Proteinase K, incubating with agitation at 60°C for 2 h. The DNA was purified using spin columns provided in the kit or using QIAquick columns (Qiagen).

A standard PCR was performed for the control samples (IgG ctr, input DNA and DNA-pol samples). Real time quantitative PCR was used to assess the fold enrichment of Runx2 binding at the rs2981578 locus (Primer table, section 2.8). A master mix was prepared as described previously and 2 µl sample (IgG or Runx2) was added, in triplicate. The programme consisted of 10 min of initial denaturation at 94°C, and 50 cycles of 20 s denaturation at 94°C, 1 min annealing and extension at 60°C. The Ct values obtained were used to evaluate the fold enrichment of Runx2 binding compared to the IgG control by calculating the ΔCt (see section 2.4.3).

<u>FOXA1 chromatin immunoprecipitation</u>

FOXA1 ChIP was carried out using an in-house protocol from Prof Ponder's lab (CRUK, Cambridge Research Institute) using 5 µg of anti-FOXA1 antibody (Ab5089, Abcam). Cells were plated in a 150 mm culture dish. After 24h, test cells were deprived of oestrogen for 3 days by replacing the media by phenol-red free DMEM (Sigma) supplemented with 5% of charcoal-stripped FBS (Gibco). The starvation medium was changed every day for three days. The starved cells were then stimulated with 100nM of β-oestradiol (Sigma) for 1 hour. The control plates were grown either in full medium or starved without oestrogen stimulation. All the cells

were then cross-linked at 37°C in 1% formaldehyde (Sigma) for 10 min. The reaction was quenched by adding 1.5 ml of 1M glycine at RT for 5 min. The cells were washed twice in cold PBS containing 1X protease Inhibitors (Roche), scraped and collected by centrifugation at 10,000x *g* for 3 min (all centrifugation steps were carried out at 4°C). The pellets were frozen on dry ice and stored at -80°C to be used for up to 1 month.

The day before immunoprecipitation, 50 µl of protein G magnetic beads (Invitrogen) were washed and blocked in 1ml PBS + 5mg/ml BSA. They were then resuspended and 300 µl of PBS/BSA and 5 µg of FOXA1 antibody was added. The beads were incubated overnight on a rotator, at 4°C.

The cell pellets were resuspended in three consecutive lysis buffers to expose the genomic DNA and sheared using a Bioruptor for 15 min (30 s sonication, 30 s rest, repeated 15 times). The debris was pelleted by centrifugation and the supernatant was added to the pre-incubated beads and incubated overnight on a rotator, at 4°C. 30 µl of input samples were removed prior to the incubation with the beads and stored at 4°C.

The magnetic beads were then washed several times in RIPA buffer (Upstate, Milipore) and the cross-linking of samples and inputs was reversed by incubation at 65°C overnight (or a minimum of 6 hours). The DNA was eluted in 70 µl of pre-warmed elution buffer (50°C) using a PCR purification kit (Qiagen).

Real time quantitative PCR was used to assess the fold enrichment of FOXA1 binding at the rs2981578 locus. Primers binding the *Greb1* promoter were used as positive control and primers recognising an intronic site of *Cyclin D1* with no FOXA1 binding site were used as negative control (Primer table, section 2.8). A master mix was prepared as described previously and 2 µl of sample or input (1:14 dilution) were added, in triplicate. The programme consisted of 15 min of initial denaturation at 95°C, and 40 cycles of 30 s denaturation at 95°C, 30 s annealing at 60°C and 30 s extension at 72°C. The Ct values obtained were used to evaluate the total amount of DNA in samples and inputs. The enrichment was normalised first to the input and then to the negative control.

2.7. Western blot analysis

2.7.1. Protein quantification

Cells plated in 6 well plates were washed with PBS and lysed in 100 µl 1X RIPA buffer, supplemented with phosphatase inhibitors cocktail II and protease inhibitors (both Calbiochem) at a 1:100 dilution. The plate was incubated for 30 min at 4°C with agitation. The cells were then scraped off the wells, transferred to 1.5 ml eppendorf tubes on ice and centrifuged at 10,000x *g* for 20 min (4°C). The supernatant was transferred to a fresh tube for protein quantification by DC protein assay (BioRad). Standards containing known concentrations (2000, 1600, 1200, 800, 400, 200, 100 and 0 µg/ml) of BSA were used in duplicate to establish a standard curve. Standards or samples (5 µl) were added to a 96 well plate, followed by 25 µl reagent A' and 200 µl reagent B. Reagent A' was prepared by mixing 20 µl reagent S and 1 ml reagent A. The plate was incubated at RT for 10 to 15 min until the colour had developed and the optical density (OD) was measured by spectrophotometry at 595 nm. The sample concentration was extrapolated from the standard curve obtained by plotting the OD against the known protein concentrations. The samples were then diluted 1:1 in NuPAGE LDS Sample buffer 4X (Invitrogen; supplemented with 100 mM DTT) and stored at -20°C until needed.

When cell number was similar across the plate, protein quantification was not required and the cells were lysed directly using NuPAGE LDS Sample buffer 2X (supplemented with 100mM DTT) for 5 min at RT. The cells were scraped off the plate, and sonicated briefly. The samples were stored at -20°C until needed.

2.7.2. SDS-PAGE

Total proteins were separated by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE). Identical volumes of protein samples (or 25 µg protein/lane when proteins were quantified) were loaded in 4-12% gradient NuPAGE gels, in NuPAGE MES running buffer (both Invitrogen) and run at 110 V, for 90 min. Proteins were transferred onto nitrocellulose membranes (Schleicher & Schuell) by electro-blotting for 3h (4 °C) at 30 V in Tris/glycine buffer with 20% methanol and transfer was confirmed with Ponceau Red staining of the

membranes. After de-staining in distilled water, membranes were incubated in blocking buffer (5% powdered milk, Marvel in 1X TBS) for 30 min at RT. Membranes were then incubated with primary antibody for 1h to overnight (RT or 4°C), washed 4 times in 1X TBST (TBS containing 0.1% Tween-20) for 5 min, and incubated with secondary antibody for 1h at RT. After additional washes the antibody was detected using ECL Chemiluminescence reagent mix (GE Healthcare) and specific bands were visualised on X-ray film (Kodak).

2.7.3. Antibodies

All antibodies were diluted in 1x TBS containing 3% v/v BSA. Anti-ERα (sc-543 HC20), anti-FGFR2 (Bek sc-122), anti-HSC70 (sc-7298) and anti-Runx2 (sc-10758) antibodies were purchased from Santa Cruz Biotechnology. Anti-Oct1 (ab51363) and anti-FOXA1 (ab5089) were from Abcam, anti-tubulin α (T5168) was from Sigma, anti-P-ERK (#9101S) was from Cell Signaling and anti-GAPDH (MAB374) was from Milipore. All polyclonal secondary antibodies coupled with horseradish peroxidase were used at a 1:1,000 dilution and were purchased from Dako.

2.8. Table of Primers

Primer name	*Sequence*	*Application*
SDM		
a1021g	CCTCTATGCAAATATGCGGTTTGGAGCAGGGAAGA	major to minor allele change
a1021g_antisense	TCTTCCCTGCTCCAAACCGCATATTTGCATAGAGG	major to minor allele change
g1021a	CCTCTATGCAAATATGCAGTTTGGAGCAGGGAAGA	minor to major allele change
g1021a_antisense	TCTTCCCTGCTCCAAACTGCATATTTGCATAGAGG	minor to major allele change
ZFN_binding_F	ATCTAAAGCTTTCCTCTGAATGCTGCTCTCGAGGATTGTGAGAGG	3 ZFN binding site mutations
ZFN_binding_R	CCTCTCACAATCCTCGAGAGCAGCATTCAGAGGAAAGCTTTAGAT	3 ZFN binding site mutations
c929t	CCTTCTGAGATCTAAAGCTTTCCTCTGAATGCTGC	1 ZFN binding site mutation
c929t_antisense	GCAGCATTCAGAGGAAAGCTTTAGATCTCAGAAGG	1 ZFN binding site mutation
t945c_g947c	TTCCCTCTGAATGCTGCTCTCGAGGATTGTGAGAGGTAG	1 ZFN binding site mutation
t945c_g947c_antisense	CTACCTCTCACAATCCTCGAGAGCAGCATTCAGAGGGAA	1 ZFN binding site mutation
c31a_t33a	TGGTCAGCTGGGGTCGTTTAAACTGCCTGGTCG	new PmeI site
c31a_t33a_antisense	CGACCAGGCAGTTTAAACGACCCCAGCTGACCA	new PmeI site
a433c_g434a_t436a	GGATGACACCGATGGTGCGGACAAATTTGTCAGTGAGAACAGTAAC	new PmeI site
a433c_g434a_t436a_antisense	GTTACTGTTCTCACTGACAA ATTTGTCCGCACCATCGGTGTCATCC	new PmeI site
c24a_a26t	GCTTGGCCTTGGGGCAACAGAACTAGTCTAGCACAAAT	new SpeI site
c24a_a26t_antisense	ATTTGTGCTAGACTAGTTCTGTTGCCCCAAGGCCAAGC	new SpeI site
t1609a	ATAAAGCATTTTTTTCACTGCATACTAGTTGTGGTTTGTCCAAACTC	new SpeI site
t1609a_antisense	GAGTTTGGACAAACCACAACTAGTATGCAGTGAAAAAAATGCTTTAT	new SpeI site

Primer name	*Sequence*	*Application*
PCR		
FGFR2IIIa F	AAGGTTTACAGCGATGCCCA	FGFR2 isoforms
FGFR2IIIb R	AGAGCCAGCACTTCTGCATT	FGFR2 isoforms
FGFR2IIIc F	GTGTTAACACCACGGACAAA	FGFR2 isoforms
FGFR2IIIc R	TGGCAGAACTGTCAACCATG	FGFR2 isoforms
ZFN_forward	GCAGAGTTTCTTGCCAGGTC	Surveyor assay
ZFN_reverse	ACATTCCACGTTAAGAGCCG	Surveyor assay
SNP_forward	TTGAGGCTCACCAAGTTCAG	sequencing of rs2981578
SNP_reverse	CTGTCCCAGAAAGCCTACAT	sequencing of rs2981579
hFGFR2_11496-11515_for	ACTGGGACTATGAAGCTGCT	cloning of repair template 2kb
hFGFR2_13647-13628_rev	CACCACAGAATTCCCTTGAG	cloning of repair template 2kb
markerSNP_forward	GATGGTGCGGAAGATTTTGT	Marker SNP sequencing
markerSNP_reverse	CCCGTATTTACTGCCGTTCT	Marker SNP sequencing
WWC2_27_F	AAATGCTACAGCAGAAGACTTCAC	Off-target sequencing
WWC2_362_R	CGAGTCTGTTAGCTCTGCTTCTT	Off-target sequencing
LFNG_72_F	GTGGGCTGGCTCTGAAGAT	Off-target sequencing
LFNG_417_R	CACACTCCCTGCACAGCTC	Off-target sequencing
DPP6_133_F	AGGATACGGGAGGATGTGCT	Off-target sequencing
DPP6_460_R	GCGAGACGCTGTATCAAAAA	Off-target sequencing
IGSF9B_59_F	CAGGGGATTAGAGCTGAGGA	Off-target sequencing
IGSF9B_436_R	AGGCATGATGGATACAGAGC	Off-target sequencing
TSPAN11_45_F	AGTCTCTCTAGGCGCAGCTC	Off-target sequencing
TSPAN11_376_R	CACAAGTGCAGAAAGGCAGA	Off-target sequencing

top1_68_F	AGGACAGGACACCACTTGCT	Off-target sequencing
top1_435_R	TAGTCTCTTGGGCAGGGCTA	Off-target sequencing
top2_70_F	GACAGGTTCCAGCAGATTCC	Off-target sequencing
top2_431_R	TGTGATTGTGGTGGCAAGTT	Off-target sequencing

Primer name	*Sequence*	*Application*
real time RT-PCR		
hGAPDH_forward	CAATGACCCCTTCATTGACC	qRT-PCR
hGAPDH_reverse	TTGATTTTGGAGGGATCTCG	qRT-PCR
HPRT_forward	CCTGCTGGATTACATCAAAGCACTG	qRT-PCR
HPRT_reverse	GTCAAGGGCATATCCTACAACAA	qRT-PCR
ERalpha_forward	GCACCCTGAAGTCTCTGGAA	qRT-PCR
ERalpha_reverse	TGGCTAAAGTGGTGCATGAT	qRT-PCR
cMyb_forward	GAAGGTCGAACAGGAAGGTTATCT	qRT-PCR
cMyb_reverse	GTAACGCTACAGGGTATGGAACA	qRT-PCR
pS2_forward	GAGAACAAGGTGATCTGCGC	qRT-PCR
pS2_reverse	TGGTATTAGGATAGAAGCACC	qRT-PCR
runx2-forward	AAGGACTTGGTGCAGAGTTC	qRT-PCR
runx2_reverse	TTACTGTCATGGCGGGTAAC	qRT-PCR
oct1_forward	CCGTCAGAAACCAGTAAACC	qRT-PCR
oct1_reverse	CCGTCAGAAACCAGTAAACC	qRT-PCR
FGFR2IIIa_for	AAGGTTTACAGCGATGCCCA	qRT-PCR
FGFR2IIIa_rev	CTGCTGAAGTCTGGCTTCTT	qRT-PCR
FGFR2IIIb_for	AAGGTTTACAGCGATGCCCA	qRT-PCR
FGFR2IIIb_rev	AGAGCCAGCACTTCTGCATT	qRT-PCR
FGFR2IIIc_for	GTGTTAACACCACGGACAAA	qRT-PCR
FGFR2IIIc_rev	TGGCAGAACTGTCAACCATG	qRT-PCR

Primer name	*Sequence*	*Application*
Taqman assay		
rs1047100	TGATGGACCCGTATTCATTCTCCAC[C/T]ACACAGGTATAATTTCCCTTGTCAG	allelic discrimination
rs755793	GTTTTTCAGCCACCGCATGGTTGGC[A/G]TTGGGTTCCCCCCGGCTGGGCAGCG	allelic discrimination
rs2981578	TTAACCTTTCTTCCCTGCTCCAAAC[C/T]GCATATTTGCATAGAGGTAAAAGCT	allelic discrimination
hsa-miR-221	AGCUACAUUGUCUGCUGGGUUUC	quantification
hsa-miR-19a	UGUGCAAAUCUAUGCAAAACUGA	quantification

Primer name	*Sequence*	*Application*
Chromatin Immunoprecipitation (ChIP)		
FGFR2_rs2981578_for	AGGTAGTGACTCTTCAAAGTTTGTTTGT	ChIP-PCR
FGFR2_rs2981578_rev	CGCCATCACAGTTAACCTTTCTTC	ChIP-PCR
Greb1_For	GAAGGGCAGAGCTGATAACG	ChIP-PCR
Greb1_Rev	GACCCAGTTGCCACACTTTT	ChIP-PCR
Kbm359-CCND1F	TGCCACACACCAGTGACTTT	ChIP-PCR
Kbm360-CCND1R	ACAGCCAGAAGCTCCAAAAA	ChIP-PCR

2.9. Table of Constructs

Vector name	*Common name*	*Antibiotic resistance*	*Size (kb)*
pJET2.1/MCF10A	MCF10A repair template	Ampicillin	5.1
pJET2.1/MCF7	MCF7 repair template	Ampicillin	5.1
pJET2.1/FGFR2b-GFP	FGFR2b -GFP	Ampicillin	8.2
pJET2.1/FGFR2b-GFP/neo	FGFR2b-GFP/neo	Ampicillin	9
PGKneotpAlox2	Neomycin selection casette (Addgene)	Ampicillin	5.6
pCAG-cre	Cre recombinase (Addgene)	Ampicillin	5.8
pmaxGFP	Transfection control (Lonza)	Kanamycin	3.5
PZFN1	ZFN1 targeting reverse strand	Kanamycin	4.2
PZFN2	ZFN2 targeting forward strand	Kanamycin	4.2
pCRII-TOPO	TOPO cloning vector	Ampicillin and Kanamycin	4

CHAPTER 3

ZFN-MEDIATED GENOME EDITING IN BREAST CANCER CELL LINES

3. ZFN-mediated genome editing in breast cancer cell lines

3.1. Introduction

In complex diseases such as breast cancer, as opposed to diseases that show Mendelian inheritance, the familial aggregation usually appears to be caused by a number of genes or genetic elements, interacting with various environmental factors (Motulsky, 2006). Genome wide association studies have identified some of these risk loci, particularly by identifying numerous SNPs associated with susceptibility to disease, altered response to drug treatment and other phenotypic variations. However the connection between most of those variants and the underlying mechanism of carcinogenesis remains unknown. Comprehensive functional validation studies at the biological level are needed to better understand the significance of these risk alleles.

3.1.1. Non-coding polymorphisms

Data from published GWAS demonstrate that 88% of disease associated SNPs occur in non-protein coding DNA, with 45% and 43% occurring in introns and intergenic regions respectively (Hindorff *et al*, 2009). Another study established that up to 71% of GWAS SNPs have a potential causative SNP overlapping a *DNase I* hypersensitive site, and 31% of loci have a candidate SNP that overlaps a binding site occupied by a transcription factor (Bernstein *et al*, 2012). The non-coding regions also include promoters and 3' untranslated regions (3'UTR) that may play a crucial role in modulating the expression of a neighbouring gene (or genes). Most of the SNPs that are localised to such regions are therefore expected to give rise to a phenotype. The high number of non-coding polymorphisms correlates with the non-homogeneous distribution of SNPs across the genome, with non-coding regions being less subject to natural selection than coding regions (Zhao *et al*, 2006), but also emphasises the increasing awareness of the importance of non-coding DNA in regulating genes. In addition to regulatory element binding sites, non-coding DNA sequences play other functions involved in transcriptional and translational regulation that can also potentially be affected by SNPs, such as the transcription of non-protein coding RNA (transfer RNA, ribosomal RNA, micro-RNA) (Jin *et al*, 2011; Ritz *et al*, 2012).

SNPs located in non-coding regions are more difficult to identify as, by definition, protein structure and/or function is unchanged. Current methods of investigating such SNPs consist of indirect *in vitro* assays such as chromatin immunoprecipitation (ChIP) and Electrophoretic Mobility Shift Assay (EMSA), that can detect the binding of proteins to a given DNA sequence. Additionally, chromatin conformation capture is used to demonstrate the physical interaction between a transcription factor binding site and a gene promoter. The putative regulatory sequences can also be cloned upstream of reporter genes, such as *Luciferase,* to assess whether the impact on the level of gene expression is SNP dependent. However, *in vitro* proof of binding does not necessarily mean that a significant phenotype will ensue.

3.1.1. Site-specific genome editing

The early stages of cancer development are accompanied by subtle phenotypic changes that can be difficult to model using *in vitro* cancer cell lines. This is because oncogenic phenotypes might be modulated by biological variation between the cells lines themselves (either caused by a prolonged culture period or inter-individual variation), concealing any SNP-specific phenotype. It is therefore essential to set up models of isogenic cell lines where the putative disease causing genetic polymorphism is the sole modified variable.

The objective of this study was to generate a panel of control and disease-associated breast cancer cell lines by editing the allele of rs2981578 to assess the importance of that SNP in the development of breast cancer, and decipher the apparent association with ER positive disease.

Genome editing techniques relying on homologous recombination have been used extensively to create knock out and knock in of genes in animal models (and plants) and cell lines to study the role of genes and/or regulatory sequences. Random transgene integrations, used extensively in microorganisms and plants in biotechnology, have the principal drawback of unpredictable gene expression due to multiple transgene copy integration and lack of control over integration sites (Dellaire and Chartrand, 1998; Conner and Jacobs, 1999). Site-specific recombination, however, is much safer but has a low efficiency. Classical targeting

approaches employ the co-integration of an antibiotic resistance cassette with the transgene to allow for positive selection of transfected clones. Such selection cassettes may be flanked by recombination sites (such as *LoxP* sites) that can be excised by recombinases (e.g. Cre) to facilitate excision of the cassette, to prevent interference with the transgene (Birling *et al*, 2009). The remaining *LoxP* site at the site of integration consist of a 34 bp sequence that could be a disadvantage when studying a region of the genome that is rich in regulatory sequence such as transcription factor binding sites and methylation sites. A key advantage of targeted genome editing using zinc finger nucleases (ZFN) is that it leaves the neighbouring DNA intact and is therefore a more suitable approach for the study of regulatory DNA. Recently, many studies have used engineered ZFNs to drive efficient genome editing in different cell types such as rat zygotes (Geurts *et al*, 2009), human embryonic stem cells (Chang and Bouhassira, 2012), human cancer cells (Gutschner *et al*, 2012) and human T cells (Geurts *et al*, 2009). Some studies use ZFN technology for gene therapy by inserting a transgene into a safe harbour gene locus, as in the study by Chang and colleagues for correction of α-thalassemia (Chang and Bouhassira, 2012). Others have attempted to modulate the response to certain anti-cancer drugs by deleting polymorphisms in the pro-apoptotic gene *BIM*, which affect the response to tyrosine kinase inhibition (Ng *et al*, 2012).

In this study, ZFN technology was used as a proof of concept to show it is possible to engineer and study functional intronic SNPs. Because of the known association between *FGFR2* SNPs and breast cancer, we determined to use rs2981578 as a model system to examine the feasibility of this approach.

3.2. Results

3.2.1. rs2981578 SNP status in a panel of breast Cancer cell lines

rs2981578 has three possible genotypes in diploid cells: (A;A), (A;G) and (G;G), where the G allele is the disease associated allele that confers an increased risk of developing ER positive breast cancer. Cloning of the rs2981578 locus and sequencing was first carried out to determine the SNP genotype of the candidate

A

Breast cell lines	rs2981578 alleles	ERα status	CN
MCF7	(A;A)	positive	2.16
T47D	(A;A)	positive	1.51
ZR-75-1	(G;G)	positive	2.20
MCF10A	(G;G)	negative	2.00
BT20	(G;G)	negative	0.87
SKBR3	(A;G)	negative	1.65
MDA-MB-453	(G;G)	negative	1.99
MDA-MB-468	(A;A)	negative	1.55
β4-1089	(G;G)	negative	N/A
H3396	(A;A)	positive	N/A
SUM159	(G;G)	negative	N/A

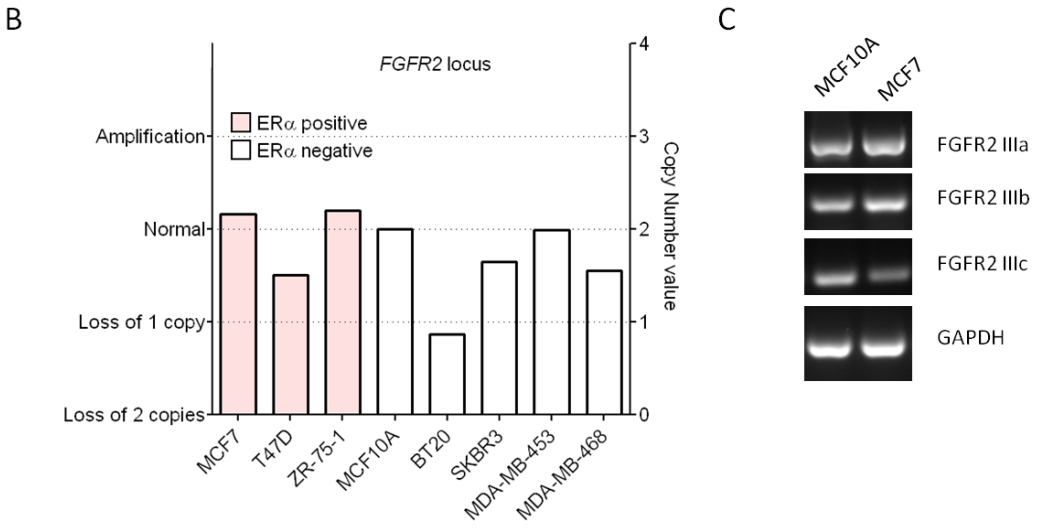


Figure 3.1: Candidate breast cancer cell lines characteristics

A) Comparison of rs2981578 SNP status, ERα status and FGFR2 copy number variation (when available) in a panel of breast cancer cell lines. Dataset from Cancer Cell Line Enclyclopedia database: DNA Copy Number Affymetrix SNP 6.0 arrays. B) Comparison of FGFR2 copy number value in a panel of ERα positive (pink) and ERα negative (white) breast cancer cell lines. CNV value of 3 is equivalent to duplication whereas a value of 1 represents the loss of one copy of a chromosome portion. CNV data were visualised using the IGV2.2 software, data obtained from http://www.broadinstitute.org/ccle (September 2012). C) RT-PCR for the different FGFR2 isoforms in MCF10A and MCF7 cell lines. The western blot is a representative image of three independent experiments (35 cycles, GAPDH serves as a control for RNA integrity).

breast cancer cell lines use for ZFN-mediated editing. Cell lines were classified depending on their ERα status, since this is the only tumour characteristic that was found to be associated with *FGFR2* dependent risk, and their respective *FGFR2* copy number, which is crucial information for genome editing of multiple alleles (Fig. 3.1A and B). Interestingly, the results indicated that all the cell lines investigated were homozygous, except the SKBR3 cell lines (Fig. 3.1A). The proportion of cell lines with the non-disease associated allele (four out of eleven A;A) was slightly lower relatively to the disease-associated allele (six out of eleven G;G). Genome editing using ZFN mainly allows the editing of one copy of the target allele at a time, but in some cases can lead to biallelic editing (12.2% vs. 2.4%) (Urnov *et al*, 2005); it is therefore probable that two rounds of ZFN specific editing would be necessary in order to change the SNP status of one diploid cell line into the other alternative homologous genotype.

Additionally, one might hypothesise that the putative phenotype of rs2981578 could be more visible in the early stage of breast cancer development, rather than at a more advanced stage, where other new mutations in oncogenes might mask any phenotypes related to the SNP; therefore candidate cell lines that possessed a relatively normal karyotype, with only two copies of chromosome 10, were favoured. Copy number variation data from the Cancer Cell Line Encyclopaedia (Affymetrix SNP6.0 Array, CCLE, Broad Institute) were used to determine whether the candidate cell lines showed *FGFR2* deletion or duplication (Fig. 3.1B). MCF10A cells, which were homozygous for the disease associated allele, were initially chosen as working models for this study, as they fit this profile and are a well characterised spontaneously immortalised breast epithelial cell line used in many studies (Wang *et al*, 2012; Scribner *et al*, 2012; Ward *et al*, 2012). The only drawback regarding the use of MCF10A cells as model was the fact that they did not express ERα (Neve *et al*, 2006). The MCF7 cell line, which is ERα positive and homozygous for the major allele of rs2981578, was another possible candidate, despite showing a small *FGFR2* amplification (Log2 (Copy Number/2)=0.113, Fig. 3.1A and B). Both MCF7 and MCF10A cells expressed FGFR2 isoform b predominantly but transcripts of the c isoform also were detected by RT-PCR (Fig. 3.1C).

3.2.1. Oestrogen receptor expression in the MCF10A cell line

MCF10A, a non-tumourigenic, spontaneously immortalised epithelial breast cell line (Soule *et al*, 1990), constitutes an attractive model for the study of cancer initiation, as the cells have not yet accumulated many genetic alterations, compared to the weakly metastatic MCF7 cell line (Neve *et al*, 2006). Indeed, their karyotypes are very different, with many abnormalities with high-level amplification in the MCF7 cells (Appendix 1). All the GWAS have consistently demonstrated that the correlation between *FGFR2* intronic SNPs and breast cancer risk was statistically more significant in ERα expressing tumours as compared to ER negative ones (Easton *et al*, 2007; Hunter *et al*, 2007; Antoniou *et al*, 2008). However, the MCF10A cell line, the favoured model cell line, is ERα negative (Neve *et al*, 2006).

Possible acquisition of ERα expression was investigated in a series of cell lines, of increasing tumourigenicity, that were derived from the original MCF10A cells after a series of mouse xenografts and induced RAS mutations (Fig. 3.2A) (Santner *et al*, 2001). Original MCF10 cells were obtained from benign breast tissue from a woman with fibrocystic disease (Soule *et al*, 1990). The series then was initiated with the mortal MCF10M and MCF10MS cells (mortal cells grown in serum-free and serum-containing media, respectively), that gave rise to the spontaneously immortalised but otherwise normal MCF10F and MCF10A lines (free-floating versus attached cells), the transformed MCF10AneoT cells, transfected with T24 *Ha-Ras*, and premalignant MCF10AT1/k.c12 cells. Additionally, fully malignant MCF10CA1 (a and h) lines were developed to complete the spectrum of progression from normal breast epithelial cells to breast cancer cells capable of metastasis. MCF10A DCIS.com is a human cell line that forms ductal carcinoma *in situ* (DCIS) when xenografted into immunodeficient mice, and was derived from premalignant MCF10AT cells (Miller *et al*, 2000). Western blot analysis for ERα in the different cell lines revealed that only a very small amount of ERα protein was

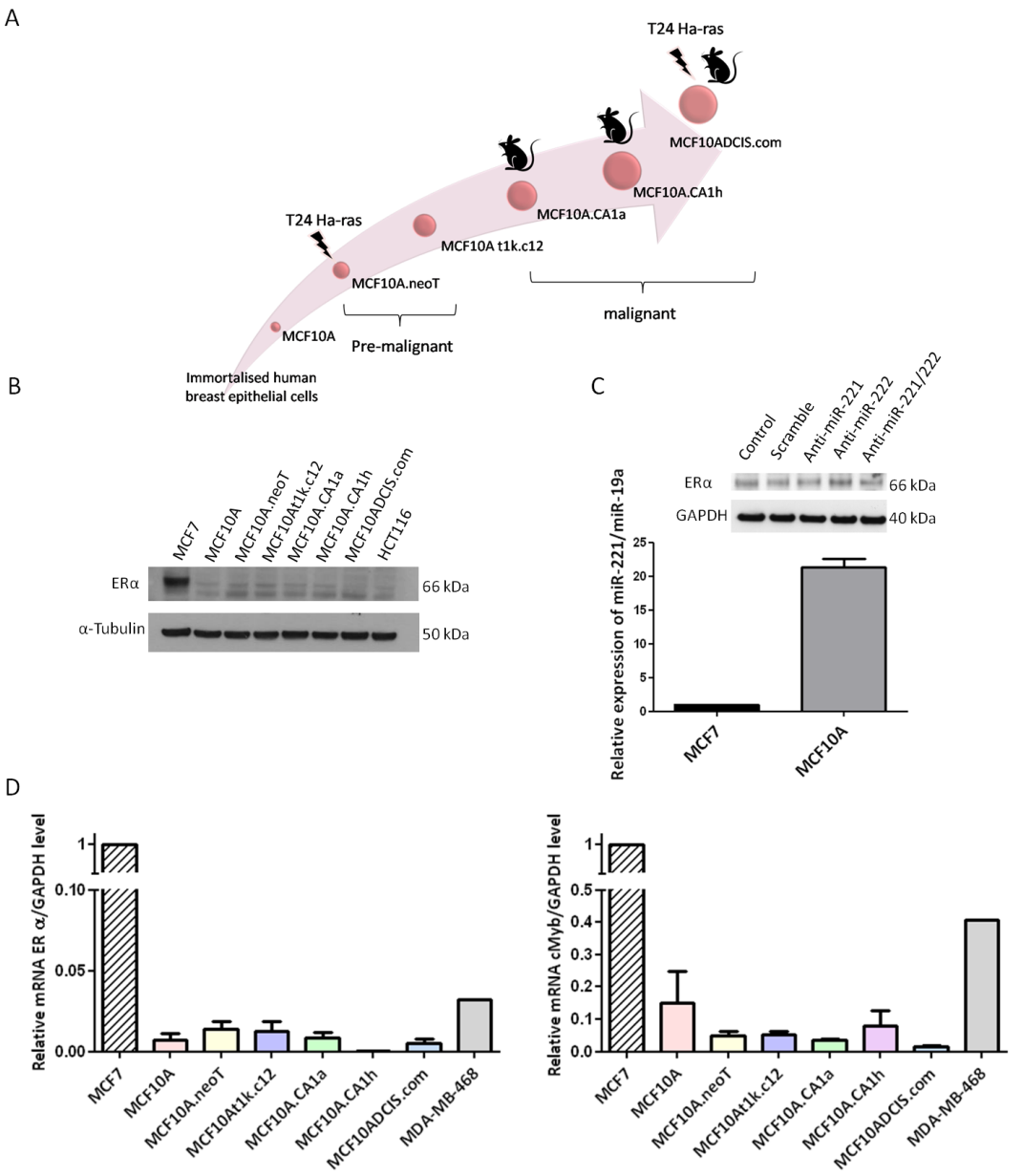


Figure 3.2: Oestrogen receptor, miR-221 and miR-222 expression in the MCF10A cell line series

A) The MCF10A cell line series. MCF10 panel of cell lines, derived from a single patient, representing sequential stages of progression (immortalised, pre-malignant and malignant) as shown by their ability to form xenograft lesions in immunodeficient mice (mouse icon)(Kadota *et al*, 2010). B) ERα expression in the MCF10A cell line series as assessed by western blot analysis. MCF7 is an ER positive cell line used for positive control. HCT116 is a colon cell line, used as negative control for ERα expression. C) Taqman assay showing miR-221 expression (normalised to miRNA-19a expression) in MCF7 and MCF10A cells, one biological experiment, three technical repeats, error bar represents SEM. Transient knock down of both miR-221 and miR-222 and the effect on ERα protein level in MCF10A cells. D) Relative mRNA expression of ERα and cMyb (an ER response gene) in the MCF10A cell series and MDA-MB-468 cells, another ER negative cell line used as control, relative to GAPDH expression. The experiment was done in triplicate, error bars represent SEM.

detected in MCF10A cells and all the other related cell lines, as compared to high levels observed in MCF7 cells, which are classified as ERα positive (Fig. 3.2B).

Interestingly, it was shown that the 221-222 microRNA cluster can target ERα mRNA and might therefore be responsible for a fraction of the ERα negative breast carcinomas or breast cancer cell lines (Zhao *et al*, 2008). Zhao *et al* reported that, although not detectable at the protein level, ERα mRNA was transcribed in MCF10A cells, indicating a post-transcriptional repression of ERα by miR-221 and miR-222. They used antagomirs, miRNAs that target endogenous miRNAs, to inhibit miR-221 and miR-222 and therefore release the ERα inhibition.The same approach was considered, to re-establish ERα protein expression in MCF10A and thus potentially create an inducible system where ERα positive and ERα negative MCF10A modified cells could be used in parallel for functional studies. The expression level of miR-221 was assessed by quantitative RT-PCR using specific Taqman probes in MCF10A and MCF7 cells. The Taqman assay used was specific for miR-221 only, however as miR-221 and miR-222 are part of the same cluster on the X chromosome and are located on the same pri-miRNA, they are co-expressed at similar levels (Yu *et al*, 2006). The results indicated that although MCF7 cells lack miR-221 expression, it was expressed at high levels in MCF10A cells and might therefore be responsible for the downregulation of ERα. However, all three attempts to knock down miR-221 and miR-222 with specific antagomirs failed to demonstrate any significant increase in ERα protein expression (Fig. 3.2C).

In an attempt to replicate the findings of the Zhao study, quantitative RT-PCR was performed to compare ERα transcript levels in the MCF10A cell series and the ER-positive cell line, MCF7. However, RT-PCR results showed very low levels of ERα mRNA in all the members of the MCF10A series (Fig. 3.2D). Similarly, low cMyb mRNA levels, a positively ERα-regulated gene (Gudas *et al*, 1995), were detected. The low level of ERα transcription is contradictory to published data (Zhao *et al*, 2008), which indicates that the MCF10A cells at our disposal must somehow differ from those used in the Zhao study. Short tandem repeats (STR) profiling of all the cell lines used by the Tumour Biology department was carried out and confirmed

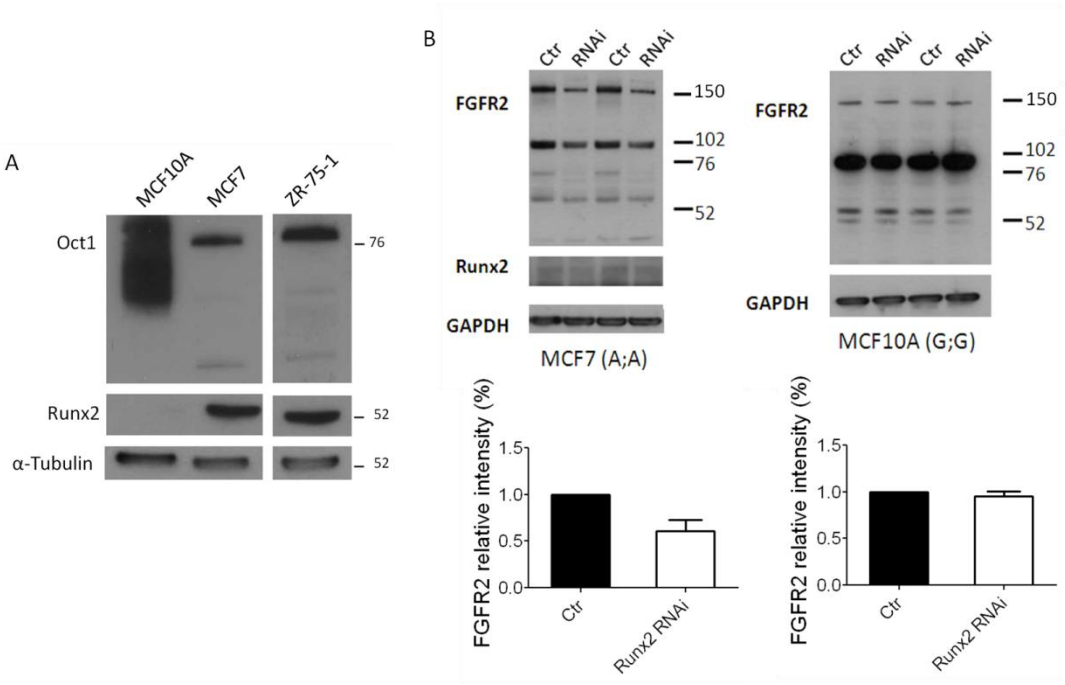


Figure 3.3: Runx2, Oct1 and FGFR2 expression

A) Western blot showing Oct1 and Runx2 protein levels in three breast cancer cell lines, α-tubulin was used as loading control. B) Transient knock down of Runx2 (shown in duplicate) using a pool of 4 different targeting siRNAs (or non-targeting control) was performed and FGFR2 protein levels were determined by Western blot in MCF10A and MCF7 cells. The two top bands are full length receptor and the lower bands represent truncated forms of FGFR2 (~60 kDa). GAPDH was used as loading control. Densitometry of FGFR2 expression relative to GAPDH, is shown below the graph and represents a total of 4 independent experiments. Error bars show SEM.

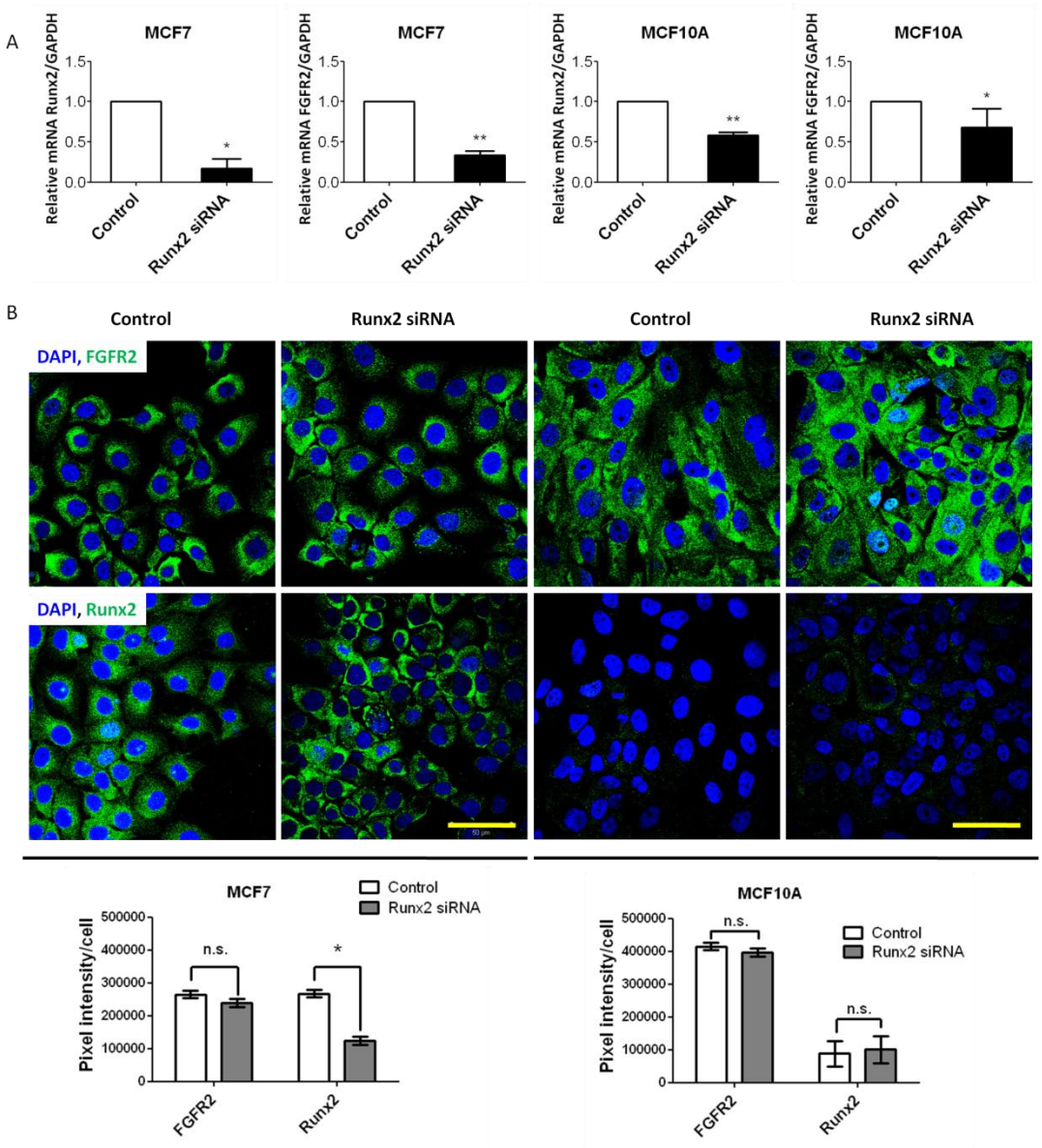


Figure 3.4: Runx2 knock down in MCF7 and MCF10A cells

A) Knock down of Runx2 in MCF7 and MCF10A cells, using a pool of four siRNAs, was assessed by quantitative RT-PCR. Paired T test was performed on triplicate experiments. B) Immunocytochemistry staining of FGFR2 and Runx2 in MCF7 and MCF10A cells 48 hours following a transient Runx2 knock down using siRNA. Scale (yellow bar) equals 50 µm. Measurement of the average green pixel intensity per cell (in 10 fields of view) was used for quantification purposes.

the identity of the MCF10A cells used in this study, indicating that we had been using the correctly designated line (Appendix 2).

This approach was therefore halted and the use of an ERα positive cell line, or one where the overexpression of ERα had been engineered was considered instead.

3.2.1. Runx2, Oct1 and FGFR2 expression

Meyer *et al.* (2008) hypothesised that Runx2 is capable of binding the disease associated allele of rs2981578 only (G;G), and acts as an enhancer of *FGFR2* transcription. Western blot analysis and immunocytochemistry were used to assess the basal expression level of transcription factors Runx2 and Oct1 in two breast cancer cell lines (Fig. 3.3A and Fig. 3.4B). MCF10A cells did not express high levels of Runx2 as seen by Western blot and immunocytochemistry, possibly consistent with reports in the literature suggesting higher Runx2 expression occurs in more metastatic samples, compared to normal tissues (Shore, 2005). Moreover, Western blotting for Oct1 only picked up a smear instead of a single band in MCF10A cells, compared to the other cell lines. A Transient knock down of Runx2, using a pool of 4 siRNA, was performed in MCF10A and MCF7 cells. Technical issues with a different Runx2 antibody lot made the protein detection by Western blot quite difficult (Fig. 3.3A and B) in MCF7 cells and no band was detected in the MCF10A blot (data not shown). Other techniques, such as real-time PCR and immunocytochemistry, were used to detect Runx2 knock down (Fig. 3.4A and B). A clear reduction in total FGFR2 protein and mRNA levels were however observed in MCF7 cells, but not in MCF10A cells (Fig. 3.3B and Fig. 3.4C). MCF7 cells are homozygous for the major allele of *FGFR2*, not associated with breast cancer risk, and therefore should not possess any Runx2 binding site at the rs2981578 locus. The decrease in FGFR2 protein level observed in Figure 3.3B thus should not be caused by the absence of the Oct1/Runx2 complex at this site. Considering that Runx2 is a key transcription factor involved in many cellular mechanisms, we decided it was unlikely that this line of investigation would be informative regarding the effect on *FGFR2* intronic SNPs independently from any other Runx2 target genes. Other approaches such as Chromatin Immunoprecipitation (ChIP) were therefore used to analyse the importance of Runx2 binding

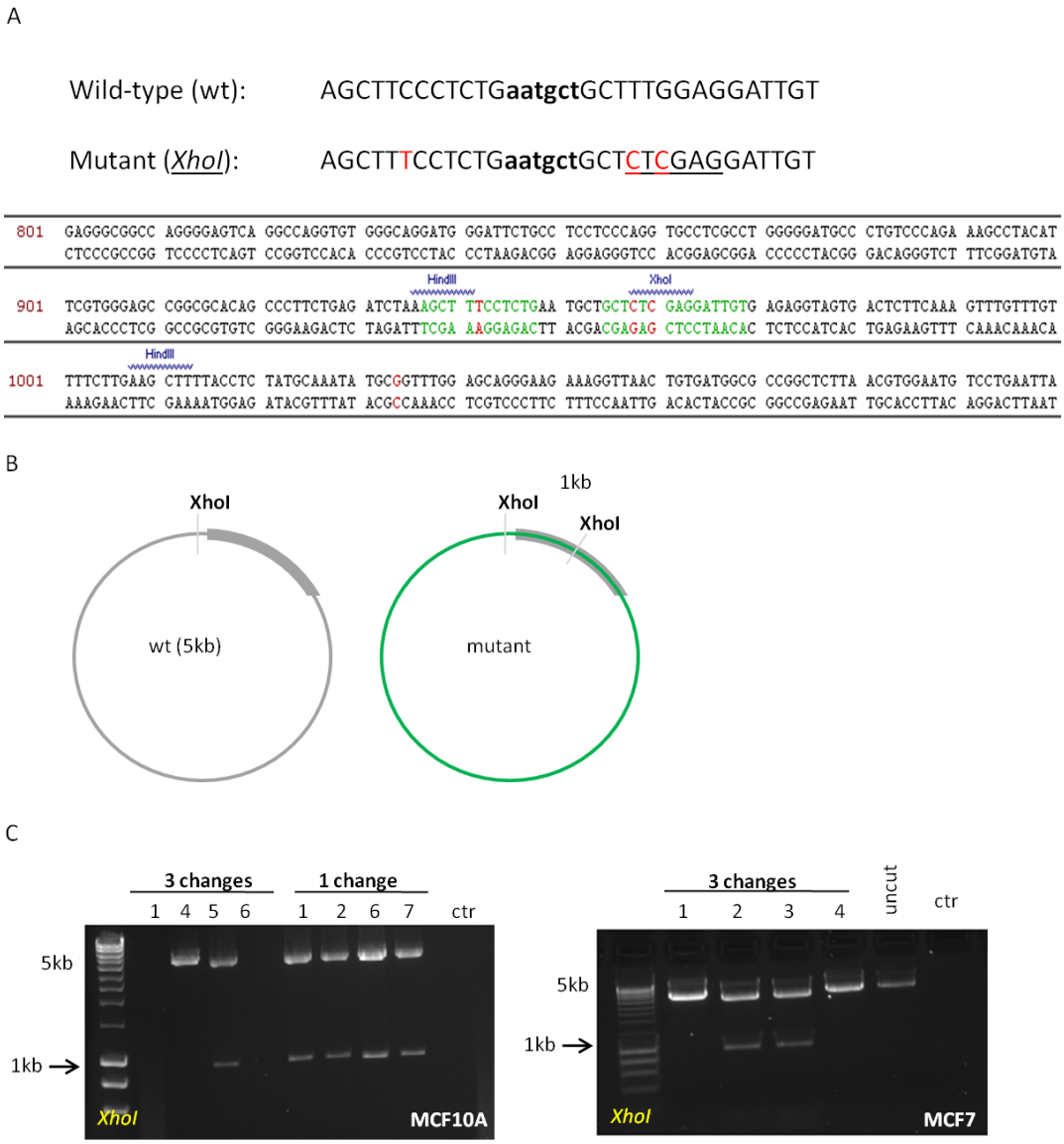


Figure 3.5: *FGFR2* donor template with modified ZFN binding sites

A) The wild-type ZFN target sequence is displayed next to the mutated sequence that impairs ZFN binding. Also shown is the region around rs2981578 (shown in red at position 1033), showing ZFN sites (green) and mutations (red). The mutated target site forms a new *XhoI* site (underlined), used for screening purposes. B) Restriction map of pJET1.2/blunt plasmid containing the 2kb insert from *FGFR2* intron two. A new *XhoI* site was created by site directed mutagenesis. C) DNA from several colonies obtained after SDM was digested with *XhoI* and resolved by electrophoresis. The mutated inserts have a different digestion profile than the wt, displaying a new band of 1 kb. Control lanes correspond to digestions with no plasmid DNA, uncut lane does not contain *XhoI* enzyme. Two alternative SDM primers (that either contained the three changes or only one at a time) were used.

in the mechanisms associated with the increased breast cancer risk.

3.2.1. Design of repair template for genome editing

Genome editing using the custom-made *FGFR2* ZFN requires a repair (or donor) template in the form of a plasmid with 1kb homology arms flanking the target SNP (a minimum of 750 bp of homology was required). The creation of *in vitro* models composed of isogenic cell lines that only differ in rs2981578 SNP status requires the absence of foreign DNA at the site of integration, excluding therefore the use of an antibiotic selection marker in the DNA repair template. However, an additional change to the ZFN binding site was recommended to reduce ZFN targeting of the repair template and to allow for more efficient additional editing rounds. Such cutting of the exogenous template would happen considerably more often considering the great abundance of the template present in the cell after transfection and consequently would reduce the chance of obtaining an allelic change. Three bases on the ZFN target site of the repair template were modified using site directed mutagenesis (SDM), creating a new *XhoI* site as a result, which was used for post-SDM screening (Fig. 3.5A). Two alternative SDM primers (that either contained the three changes or only one at a time) were used (Primer table, section 2.8). The multiple changes approach was successful as three plasmids (one from MCF10A cells and two from MCF7 cells) contained the new *XhoI* site as shown by restriction enzyme digestion (Fig. 3.5C).

Plasmid number 5 in MCF10A cells and number 2 in MCF7 cells were used to create the repair templates. An additional round of SDM was used to edit the rs2981578 locus, either replacing the major allele in MCF7 (A;A) or the minor allele in MCF10A (G;G). In the absence of restriction sites, the SNP change on the repair template was assessed by sequencing. An additional repair template for MCF7 cells was created that contained the change in SNP status only, without the modified ZFN binding site, to be tested for genome editing alongside the original template.

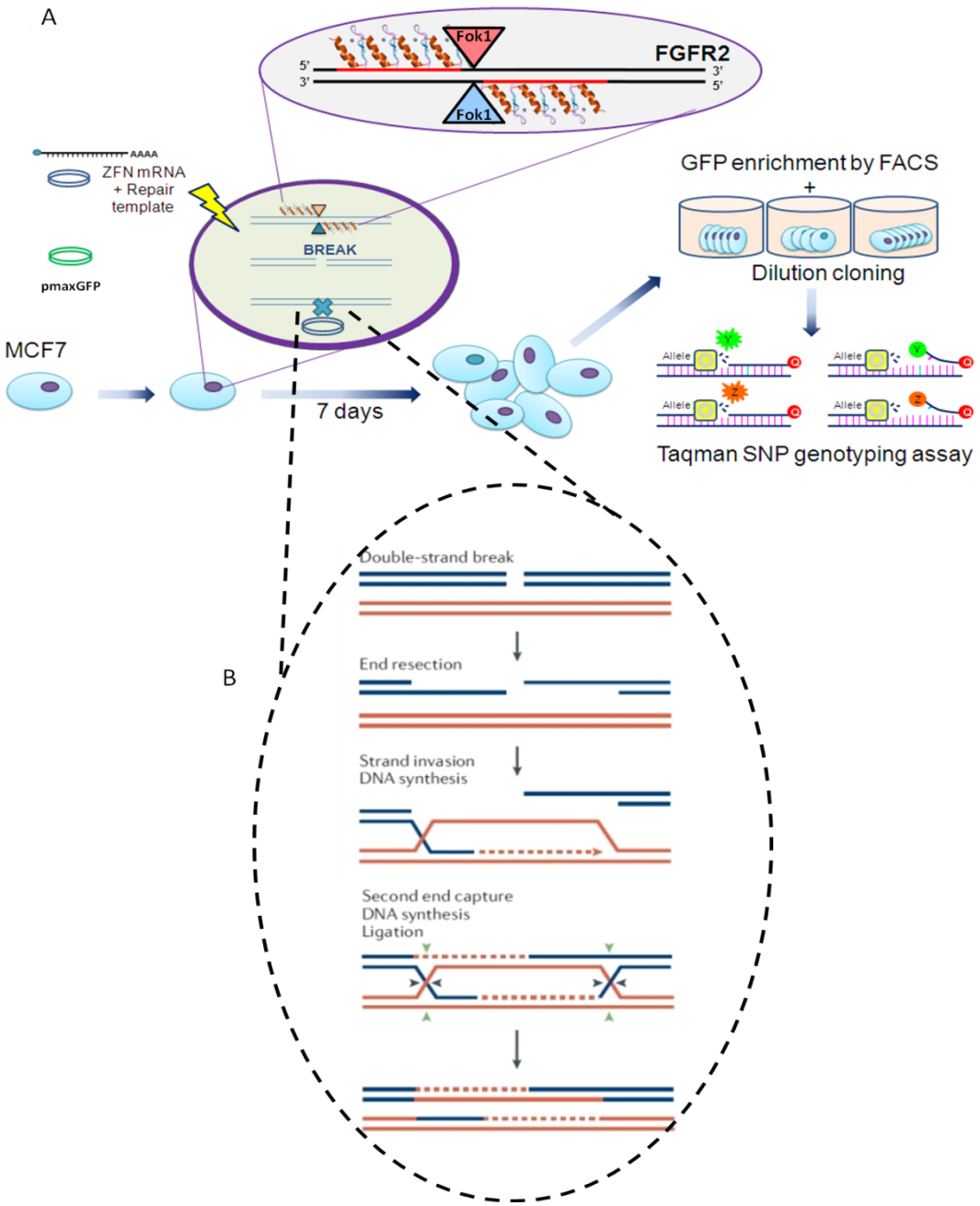


Figure 3.6: Workflow of the ZFN-mediated genome editing process in the MCF7 cells

A) MCF7 cells were transfected using Amaxa nucleofection with *FGFR2* ZFN mRNAs, repair templates containing the rs2981578 risk allele and a pmaxGFP plasmid. At the *FGFR2* SNP locus, the ZFN pairs, previously translated by the cells' own translation machinery, was able to induce DNA double stranded break (DSB) and direct the cell DNA damage repair sytem to this specific locus. The DNA damage was repaired by homologous recombination using the exogenous repair template, present in excess after transfection. An heterozygous population of MCF7 cells were therefore obtained. GFP enrichement by FACS was performed and clonal populations were screened using a SNP genotyping Taqman assay specific for rs2981578. B) Homologous recombination is initiated by resection of DSB to provide 3' single-stranded DNA (ssDNA) overhangs. Strand invasion by the 3' ssDNA overhangs into the repair template (red) is followed by DNA synthesis at the invading end. The second DSB end can be captured to form an intermediate with two Holliday junctions (HJs, green arrow heads). The final structure is resolved by DNA synthesis and ligation (modified from Sung et al, 2006).

3.2.1. ZFN editing in breast cancer cell lines

The CompoZr™ custom made *FGFR2* ZFNs, purchased from Sigma-Aldrich, were designed following a screening of the region surrounding the rs2981578 locus, to validate the optimum cutting site. The algorithm used by Sigma, constantly updated by Sangamo, has been designed to yield obligate-heterodimer pairs of ZFNs with high binding affinity to their specific target site. Several ZFN pairs were designed and the best performing ones were then assembled and validated biologically in HCT116 cells, a human colon carcinoma cell line (cell line of choice used by Sigma/Sangamo during the ZFN quality control process) (Park *et al*, 1987). Custom forward and reverse primers were provided for sequencing (Appendix 3 and 5).

The workflow of the ZFN-mediated genome editing process is summarised in Figure 3.6A. ZFNs are synthetic modular molecules made from the fusion of zinc-finger DNA-binding domains to the catalytic domain of the endonuclease *FokI* (Urnov *et al*, 2005). ZFNs create a double stranded break at the *FGFR2* SNP locus, thereby directing the DNA repair machinery to this site. For genome editing, the sister chromatid that would, in the normal context of homologous recombination (Fig. 3.6B), be used as a template for DNA damage repair, is bypassed in favour of a synthetic repair template (Fig. 3.5) that contains the alternative allele and is present in vast molar excess. The efficiency of this targeted recombination is far higher than normal homologous recombination. However, in the absence of a selectable marker, clonal cell populations are used for screening, using an allele specific Taqman SNP genotyping assay.

Several approaches to transfect ZFNs into a cell were tested, for transfection either in the form of mRNA or as an expression vector. The most efficient approach for the transfection of the *FGFR2* ZFNs was electroporation of the nucleic acids as mRNA. However, the amount of cell death using that method was higher than expected and could only be used with the MCF7 cells. The high percentage of cell death in MCF10A cells (data not shown) impaired their transfection efficiency. Therefore the less toxic method of lipid-based transfection, using Lipofectamine 2000 with serum-free medium supplemented with Glutamax, was favoured.

3.2.2. Assessment of DNA cleavage by Surveyor assay

To assess the cutting ability of the ZFN pairs, 2 µg of ZFN mRNAs were transfected in each cell line. Transfection efficiency, using pmaxGFP (Lonza), was estimated to be between 30% and 50% in MCF10A cells (data not shown) and 60 to 70% in MCF7 cells (Fig. 3.7A). ZFN cutting efficiency was assessed using the Surveyor™ assay (Transgenomic) (Fig. 2.3). Cells transfected with either ZFN mRNA or pmaxGFP were harvested for genomic DNA preparation 24 hours after transfection. In the absence of repair template, double-stranded breaks are not resolved by homologous recombination but by the non-homologous end joining pathway instead, which leads to small errors such as indels (insertions and/or deletions) (Lieber, 2008). The Surveyor mutation detection assay is based on the *Cel-I* endonuclease, derived from celery, that specifically cleaves small mismatches caused by SNPs, small insertions or deletions and can therefore be used to detect those potential errors and to estimate the frequency of double stranded breaks in control transfected cells relative to those transfected with the custom ZFN pair (Fig. 2.3). A 333 bp fragment containing the ZFN target site was amplified by PCR. The products were denatured and re-annealed, allowing the formation of homoduplexes and heteroduplexes between the DNA molecules that had been cut and repaired and those which were not cut. The percentage of cleaved products compared to the intact DNA was visualised after digestion with *Cel-I* (Fig. 3.7B). A new band, reflecting ZFN-mediated cleavage is present in the ZFN transfected lane and absent in the GFP control lane. The ratio of band intensity should, in theory, represent the ZFN cutting efficiency but this is difficult to quantify with such faint secondary bands. The presence of such bands was an indication that cleavage had occurred and constituted validation for the ZFN mRNA. An estimate of the ZFN cutting efficiency is important to determine the number of clones required for screening; however it was assumed, from the Surveyor assay data, that the efficiency was rather low and that screening of around 100 clones was the best strategy for this study.

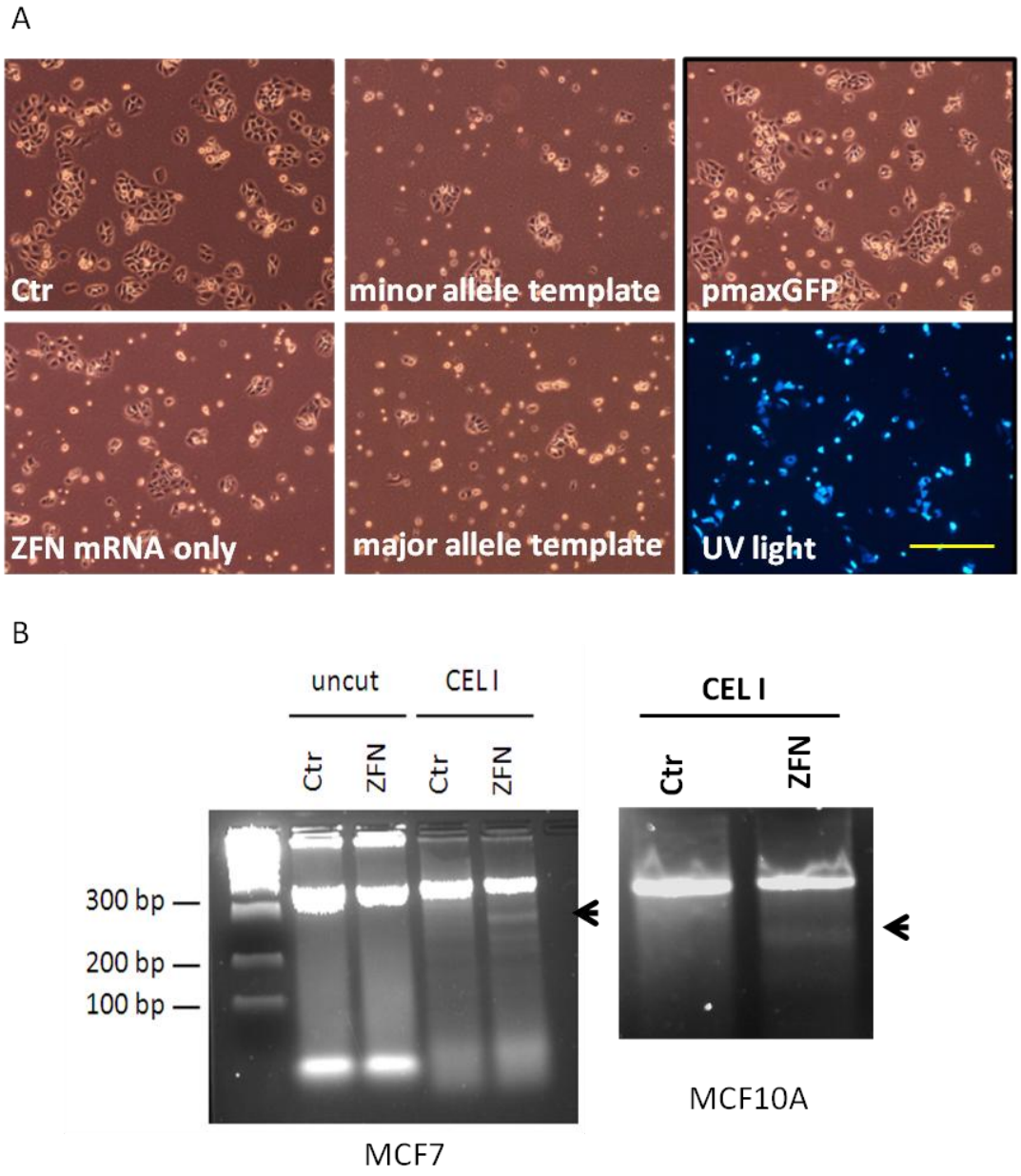


Figure 3.7: Transfection efficiency and Surveyor assay

A) Transfection of pmaxGFP shows a transfection efficiency of about 50% of the cells as assessed by the number of GFP positive cells under a UV microscope. ZFN mRNA transfection without the DNA repair template was performed to be used for the Surveyor assay to control the cutting efficiency of the ZFN pair. Scale bar represent 200 µm. B) MCF7 and MCF10A cells were used for Surveyor assay. The samples were resolved on 10% non-denaturing polyacrylamide gels. A new band (black arrow head) was present in the ZFN mRNA samples compared to the GFP control, reflecting the modification induced by the ZFN cutting of genomic DNA.

3.2.3. Single cell cloning and screening

a. MCF10A cell line

The slow growing MCF10A cells performed very poorly in initiating colonies from a single cell in culture. Feeder cells (NIH 3T3 fibroblasts previously treated with mitomycin C) were therefore used to support the growth of the clonal populations. MCF10A cells are strongly adherent cells that often display plasma membrane lamellipodia and filopodia, and were therefore more difficult to detach from their substrate, making the colony picking method, used for MCF7 cells, poorly suited to this cell line.

As an initial attempt, 48 MCF10A single cell colonies (in duplicate) were screened using SNP genotyping Taqman assay for rs2981578. The number of cells present in each well was highly variable, as not all colonies grew at the same rate. The Taqman assay was therefore performed with different DNA concentrations. Interestingly, the samples with the highest DNA concentrations did not give a strong signal (Fig. 3.8B). On average, a DNA concentration around 10 ng/µl was optimal for the assay conditions.

Overall, the majority of the colonies, from the first screening attempt, were homozygous for the disease associated SNP allele (G), with the exception of three samples that contained the modified allele (A) (Fig. 3.8C, blue circles). Among these three, two originated from the same colony in duplicate. The third modified colony was not selected as each of its duplicates gave a different result, potentially indicating a problem of contamination with gDNA from some other wells.

Additionally, the only modified clone obtained failed to grow further in culture, displaying quiescent cell features. The addition of NIH 3T3 fibroblasts, treated with mitomycin C, triggered an increase in proliferation, which stopped shortly after the fibroblasts died. After no visible change in proliferation within 3 weeks, the culture was terminated. In a new attempt to modify the SNP status of MCF10A cells, the ZFN transfection and screening was repeated. This time the cells were transfected with either a minor or a major DNA repair template in order to obtain a control cell line with only the ZFN binding site change. The second attempt of

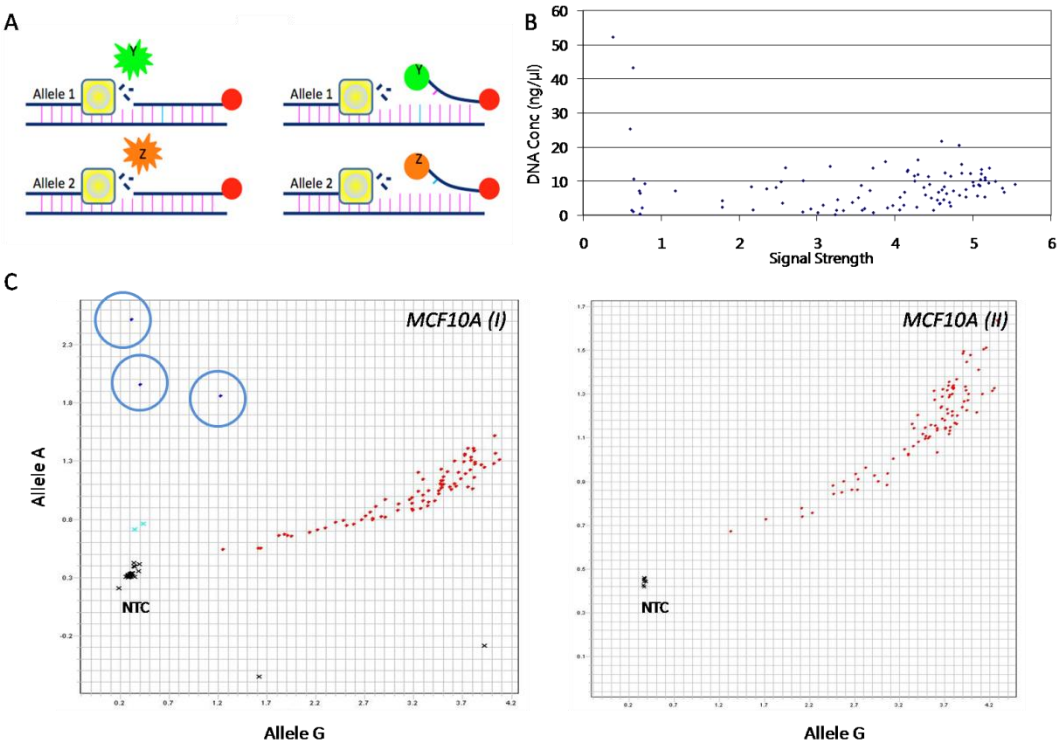


Figure 3.8: ZFN-mediated genome editing of MCF10A cell line

A) Allelic discrimination is achieved by the selective annealing of TaqMan probes. Only a matching probe is degraded by the DNA polymerase (yellow) exonuclease activity, releasing the fluorophore and allowing emission of fluorescence, as shown in the left hand sideof the figure. B) DNA concentration against Total Signal Strength. The samples with a concentration above 25ng/µl failed to yield genotyping results. C) Allelic discrimination for rs2981578 presented as a plot of fluorescence signal strength for allele (A) against allele (G). The two graphs represent the two attempts using MCF10A cells transfected with ZFN and major allele repair template. Three heterozygous clones (including two duplicates), containing one or two allele A were obtained at the first attempt (blue circles), unlike the second time, where all clones screened were wild-type (G;G). Non template control (NTC) was used to determine the basal level of background fluorescence.

modifying the SNP status of the MCF10A cells failed, as demonstrated by the allelic discrimination plot (Fig. 3.8C (right)) where all of the 68 clones screened displayed the wild-type allele.

b. MCF7 cell line

Two alternative approaches were adopted for single cell cloning of MCF7 cells. The first one consisted of creating serial dilutions from a concentrated cell suspension until reaching a concentration of one cell per well in a 96 well plate (Fig. 2.1). This strategy yielded only a small number of single cell colonies per plate (between 1 to 4 single-cell colonies). The MCF7 cells also struggled to grow at very low cell density but formed, relatively readily, loosely attached, self-contained colonies within a week in a large culture plate. The colonies were visible by eye, thus making direct colony picking, or the use of cloning rings, an attractive solution.

After ZFN delivery, and addition of a transiently expressed GFP expressing plasmid, the cells were subject to GFP enrichment by FACS, followed by single colony picking: 93 colonies were then screened using Taqman assay for rs2981578 (Fig. 3.9A). The result showed three heterozygous (A;G) MCF7 clones (3.2% efficiency), in the following location of the plate: A8 (Clone Het 1), C4 (Clone Het 2) and G11 (Clone Het 3). Unmodified MCF7 clones, homozygous for the wild-type allele (A;A) were chosen as control clones: E4 (Clone Ctr 1), F4 (Clone Ctr 2) and E11 (Clone Ctr 3). The names Ctr 1 to 3 and Het 1 to 3 are used to designate these MCF7 clones hereafter. The genotypes of the six clones were confirmed by direct sequencing and Taqman assay (Fig. 3.9B).

c. Biallelic change in MCF7 heterozygous clone

Another round of ZFN-mediated genome editing was carried out in the clone Het 2 in order to obtain a MCF7 clone that carry the genotype (G;G) for rs2981578. After screening 72 clones by Taqman genotyping assay, none of the clones carried the second allele modification. Recent studies suggested that short single-stranded oligodeoxynucleotides (ssODNs), instead of double stranded donor templates, could be used as an alternative DSB repair template for ZFN-driven genome editing

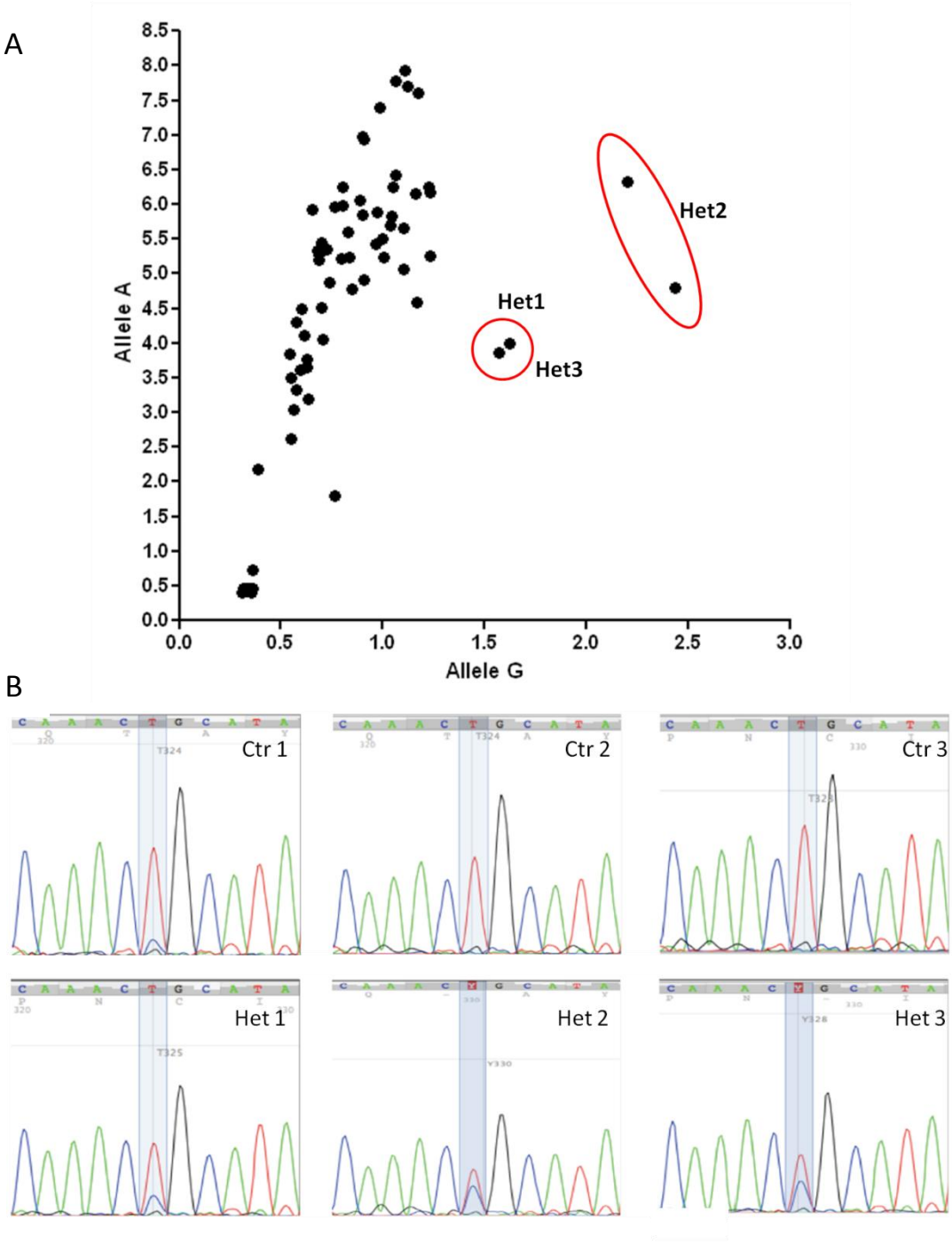


Figure 3.9: Sequencing of rs2981578 in ZFN-edited MCF7 clones

A) Allelic discrimination for rs2981578 presented as a plot of fluorescence signal strength for allele (A) against allele (G). Heterozygous clones are circled in red. Non template control (NTC) was used to determine the basal level of background fluorescence. B) Cycle sequencing performed from PCR products of the three heterozygous and three wild-type clones. The wild-type genotype of MCF7 (A;A) was identical to the non-modified controls, whereas the heterozygous clone sequencing traces displayed two overlapping peaks for nucleotide A and G.

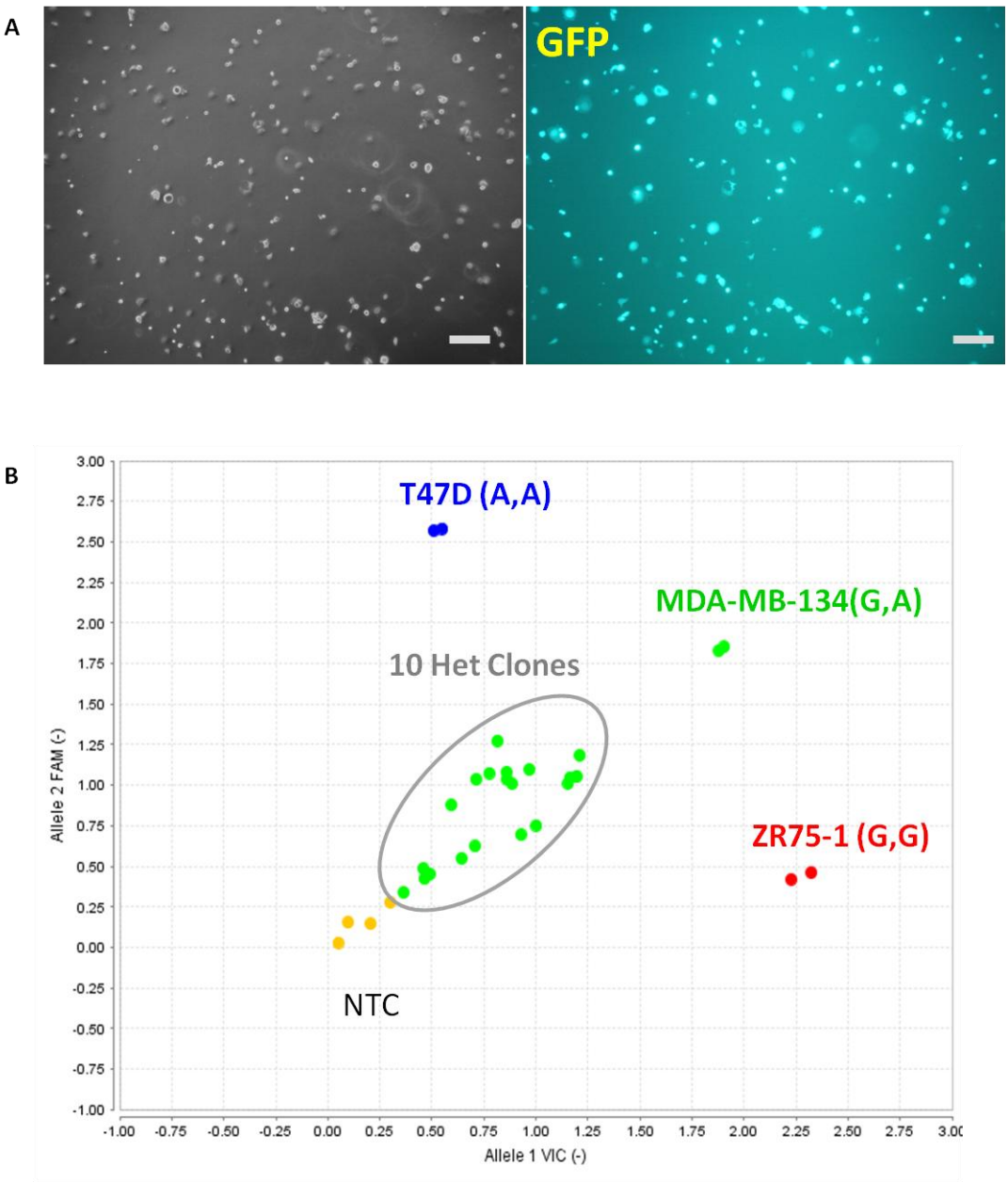


Figure 3.10: Biallelic change in MCF7 clone (Het 2)

Sequence of the single stranded oligodeoxynucleotide containing the rs2981578 risk allele used a repair template for ZFN-mediated genome editing. A) GFP expression in Het 2 cells 48 hours post transfection, showing a 90% transfection efficiency, scale bar =50 microns. B) Taqman assay showing the genotyping of the 17 clones screened for biallelic change. The GFP-positive cells were FACS sorted 48 hours post-transfection. T47D, MDA-MB-134 and ZR75-1 genomic DNA were used as controls for the different rs2981578 genotypes. Allele 1(VIC) correspond to the G allele, and Allele 2 (FAM) correspond to the A allele. NTC is the non-template control.

(Soldner *et al*, 2011). A 137 base ssODN was synthesised to be used as a repair template for genome editing using the ZFNs. The ssODN, the ZFN pair (mRNA) and a GFP vector (pmaxGFP) were electroporated in Het 2 cells using Nucleofection. The GFP expression 48 hours post transfection was used as an indicator for transfection efficiency (Fig. 3.10A) and used for single cell sorting using flow cytometry. 288 single cells were distributed in 96 wells plates containing NIH 3T3 fibroblasts (treated with mitomycin C). After 10 days, only 17 clones had proliferated sufficiently to be used for SNP genotyping screening. T47D (A,A), MDA-MB-134 (G,A) and ZR75-1 (G,G) were used as controls for each SNP genotype. The results showed that none of the clones were successfully modified to homozygosity (Fig. 3.10B). The number of viable clones obtained was insufficient to obtain a modified clone.

3.3. Discussion

GWAS have identified a haplotype in linkage disequilibrium in the large second intron of *FGFR2* associated with risk of developing ER positive breast cancer. Meyer *et al* (2008) observed that FGFR2 expression was significantly higher in patients with the disease-associated alleles of the intronic haplotype and identified rs2981578 as the putative functional polymorphism. The objective of this study was to generate a panel of isogenic breast cancer cell lines differing only in the genotype of the disease-associated allele of rs2981578, to further investigate the potential mechanism by which this allele affect breast cancer susceptibility.

None of the four ER positive cell lines screened were heterozygous for rs2981578 and only one, ZR-75-1, possessed the breast cancer risk allele. The study of risk variants that only confer a small risk increase requires the need for diploid cell lines with few oncogenic mutations. The MCF10A cell line constituted an ideal candidate, except for its lack of ERα expression, and was chosen for genome editing in order to engineer its genotype to the non-disease associated allele. The MCF7 cell line, an ER positive cell line bearing the opposite genotype, was also chosen. The problem caused by the lack of ERα expression in MCF10A cells appeared reversible in the light of a study on the role of miRNAs in cancer progression (Zhao *et al*, 2008). Indeed, miR221 and miR222 are co-expressed

miRNAs associated with the ER negative, aggressive basal-like breast cancer subtype and have been found to mediate metastasis through the regulation of epithelial to mesenchymal transition (EMT) (Stinson *et al*, 2011). Zhao *et al* (2008) observed that miR221 and miR222 were overexpressed in some ER negative breast cancers, but not in most ER positive cases, consistent with our real time PCR results comparing miR221 expression in MCF10A and MCF7 cells (Fig. 3.2C). They then demonstrated that the expression of miR221 and miR222 synthetic mimetics in MCF7 and T47D cells (both ER positive) was capable of blocking ERα expression and that, conversely, the inhibition of these miRNAs in MCF10A cells restored ERα protein expression. The results showed that the MCF10A cell line did not express ERα mRNA and therefore the knock down of miR221 and miR222 was not successful in re-establishing ERα in those cells (Fig. 3.2D). The data strongly support the idea that the oestrogen receptor alpha is not transcribed in MCF10A cells because this transcript is also absent in all MCF10A derived cell lines (Fig. 3.2D), which is further supported by a published study showing very low levels of ERα mRNA in the MCF10A cell line series (Fu *et al*, 2010).

Meyer and colleagues reported that a new binding site for Runx2, an important transcription factor implicated in osteoblast differentiation, is created by the presence of the disease-associated allele of rs2981578 (Meyer *et al*, 2008). It was hypothesised that Runx2 was acting as an enhancer of *FGFR2* expression, in association with the transcription factors Oct1 and C/EBPβ. In preliminary studies, we aimed to assess the expression of those proteins in the candidate cell lines. Runx2 was expressed in MCF7 cells but not detected in MCF10A cells by Western blot, consistent with the progressive increase in Runx2 expression in more metastatic cells compared to normal cells (Shore, 2005). Small amounts were, however, detected using real time RT-PCR and immunocytochemistry. Runx2 knock down using siRNA in MCF7 cells led to a decrease in FGFR2 protein levels, as previously reported (Zhu *et al*, 2009) but no such decrease was observed in MCF10A cells. MCF7 cells are homozygous for the non-disease associated allele of *FGFR2* and should therefore not possess any Runx2 binding site at the rs2981578 locus. The decrease in FGFR2 protein level observed (Fig. 3.3B) should, consequently, not be caused by the absence of the Oct1/Runx2 complex at this

site. Considering that Runx2 is a key transcription factor involved in many cellular mechanisms, it is unlikely that total Runx2 knock down would be informative regarding the effect on *FGFR2* intronic SNPs. The endogenous low levels of Runx2 might be responsible for the lack of change in *FGFR2* expression in MCF10A cells, making the knock down inconsequential.

The DNA repair templates were designed using genomic DNA isolated from MCF10A and MCF7 cells: one kb of DNA sequence each side of rs2981578 and the SNP status was modified, along with three bases located in the ZFN binding site. The ZFN and the appropriate DNA template were then tested in MCF10A and MCF7 cells.

The *FGFR2* ZFN cutting efficiency was lower than previously anticipated, reaching only 3% in both cell lines and only producing monoallelic changes. In a study reporting the use of ZFNs, the efficiency of the ZFN pair used for correction of point mutations at the endogenous locus of the *interleukin-2 receptor-γ* (*IL2RG*) gene in K562 cells reached 20% in the absence of any selection marker (Urnov *et al*, 2005). Importantly, 8% of the single cell derived clones obtained showed biallelic modifications. Comparable gene correction frequency was also observed, in the same study, at this locus in human CD4+ T cells. The low efficiency encountered with the *FGFR2* ZFNs might be caused by the spatial constraint of the ZFN cutting site (100 bp away from rs2981578) in the SNP region as opposed to other gene areas that are less GC rich (54.62% in the SNP region compared to the average 46.98% in the second *FGFR2* intron). Another explanation may be the low levels of homologous recombination in the MCF10A and MCF7 cell lines (as assessed by Surveyor assay). Moreover, the ZFN mRNA synthesis process was about 50% less efficient than reported by Sigma, making the total volume of ZFN mRNA transfected higher than recommended, possibly impairing the transfection efficiency.

Transient cell hypothermia has been shown to further increase ZFN-driven double strand break frequency in transformed and primary cells by two to five fold (Doyon *et al*, 2010) but did not improve the efficiency of the *FGFR2* ZFN pair. The low cell proliferation rate of MCF10A cells added another level of difficulty in terms of

single cell clone formation for screening and use of this cell line was stopped. It is important to note that a recent study, that reported using ZFNs for changes of point mutations associated with early onset of Parkinson's disease, had very low editing efficiency (4 out of 480, and 1 out of 240 clones screened had the desired nucleotide change) and overcame this problem by co-transfecting a GFP plasmid transiently, to enable GFP-positive FACS sorting of single cells (Soldner *et al*, 2011). Using this additional step of GFP-positive cell enrichment by FACS was successful, and the SNP rs2981578 was eventually modified in the ER positive MCF7 cell line only (Fig. 3.1A), resulting in a set of six clones: three controls and three heterozygous clones (Fig. 3.9).

The combination of high-fidelity DNA recognition by the ZFN pairs and homology-directed repair of ZFN-induced double-strand breaks has allowed this technology to be used for the comprehensive functional study of risk polymorphisms. The results presented here establish a proof of concept for the permanent modification of intronic SNPs in cell line models, but also highlight the inherent difficulties of using low efficient ZFNs for single nucleotide modification without the use of a selection marker.

CHAPTER 4

MCF7 CLONE CHARACTERISATION

4. MCF7 clone characterisation

4.1. Introduction

Mutations and polymorphisms in various genes and their regulatory elements are implicated in tumour initiation, progression and drug resistance (Cimoli *et al*, 2004; Sur *et al*, 2009). The understanding of the consequences of such genetic changes relies on the availability of genetic and cancer models. Site specific genome editing that does not leave any scar on the DNA, and therefore no genomic alteration, was achieved using ZFN and homologous recombination, resulting in a panel of control and disease-related breast cancer cell lines. The panel is composed of three MCF7-derived clones heterozygous for rs2981578, and three MCF7-derived wild-type controls that lack the disease associated allele of the SNP. The homozygous clones containing two copies of the risk allele could not be established after ZFN editing was repeated (one attempt with the normal repair template and another with the ssODN template). Additionally, several attemps to produce the same array of clones in MCF10A cells failed due to the difficulty of maintaining single cell cultures with this cell line.

rs28981578 is a polymorphism contained in the *FGFR2* intronic haplotype that has been associated with increased risk of ER positive breast cancer. However, one copy of the risk allele confers a 1.2 increase in risk for breast cancer development, and this figure goes up to 1.64 for individuals that carry two copies of the allele (Fig. 1.5) (Easton *et al*, 2007). It was hypothesised that the risk was mediated via an upregulation of *FGFR2*, which acts as an oncogene in breast cancer. FGFR2 signalling, principally through the MAP kinase pathway, is implicated in many cellular mechanisms including proliferation, migration, and survival (Turner and Grose, 2010). In order to detect the impact of the single nucleotide change that was engineered in MCF7 cells, heterozygous MCF7 clones were compared to their control counterparts in a series of *in vitro* assays. The mechanism of action of rs2981578 in mediating the breast cancer risk was also investigated using Chromatin immunoprecipitation followed by SNP genotyping Taqman assay in order to evaluate the degree of allele specific binding (ASB) in the heterozygous clones. Overall, this chapter is dedicated to characterising the MCF7 clones

obtained and assessing their capacity to become a model for the study of the intronic *FGFR2* SNP in breast cancer.

4.1. Results

4.1.1. Assessment of the off-target effect of *FGFR2* ZFNs

A potential limitation of ZFN-mediated genome editing is the induction of DNA strand breaks at sequences other than the intended target site. To examine off-target effects in the heterozygous clones, DNA binding affinity for other, less specific, genomic target sites was examined using the 'ZFN-site' database (Cradick *et al*, 2011) to allow the identification of the most probable off-target cleavage sites. Genotyping of the top seven off-target sites was performed to reveal any potential modification or deletion, caused by non-homologous end joining (NHEJ). The results from the software algorithm revealed no other perfect match other than *FGFR2* (Fig. 4.1A). All the other potential non-specific binding regions only allow for a five nucleotide long spacer region, compared to six nucleotides at the original *FGFR2* site, making the binding and subsequent cutting very unlikely. The top two hits were located in an intergenic region of the genome, and the following four hits were localised in non-coding regions (introns or promoters). The only site detected in a coding region was located in the membrane scaffolding protein Tetraspanin 11 (*TSPAN11*). Typically, off-target effects after a ZFN-mediated cut are repaired by NHEJ and can be visualised as 8 or 9 base pair deletions (Hockemeyer *et al*, 2009). Sequencing of these regions was sometimes difficult due to the high GC content (*LFNG* and the second intergenic hit for instance), but no such deletions were detected in any of the loci investigated, except for *IGSF9B* in Ctr 3 clone, the sequencing trace of which was not of sufficient quality to yield reliable base pair calling (Fig. 4.1B) (Appendix 7 and 8). Overall, this indicated that the *FGFR2* ZFN pairs were highly specific and did not appear to cause any off-target modification of the genome apart from the rs2981578 allele change. Additionally, unbiased screening, such as genome-wide integration site analysis (using integrase-defective lentiviral vectors) might be necessary to address this question (Gabriel *et al*, 2011). Furthermore, one of the hallmarks of cancer cells is

A

Sequences	Mismatches	Target genes
AGCTTCCCTCTGAATGCTGCTTTGGAGGATTGT NNNNNN	0	*FGFR2 (intron)*
AACTTCCCTCAGGACCCAAGAGGGAACCT G T NNNNNC G	4	*intergenic*
AGCTTCCGGCTGACCAACAGAGAGAAGCA CT NNNNN G T	4	*intergenic*
AGCTTTTCTCTGCAGTCCAGTGGGAAGCA CC NNNNN A T	4	*WWC2 (intron)*
CACTTCCCTCTGGGTTTCAGAGGGCAGCT AG NNNNN A	3	*LFNG (intron)*
AGCATCTCTCTGAATTAGAGAGGGAATCT T C NNNNNC G	4	*DPP6 (intron)*
AGCTTCCCTCTCTAGGGCAGAGGGAAGGC GNNNNN CT	3	*IGSF9B (intron)*
AGCTTCCCCCGGGCGGGCAGAGGGAAGCC T T NNNNN T	3	*TSPAN11 (exon)*

B

		clones					
Off-targets	Position	Ctr 1	Ctr 2	Ctr 3	Het 1	Het 2	Het 3
intergenic 1	Chr 2[46429876..46429904]	✓	✓	✓	✓	✓	✓
intergenic 2	Chr 2[105944949..105944977]	✓	✓	✓	✓	✓	✓
WWC2 (intron)	Chr4[184191133..184191161]	✓	✓	✓	✓	✓	✓
LFNG (intron)	Chr 7[2565539..2565567]	✓	✓	✓	✓	✓	✓
DPP6 (intron)	Chr 7[154445995..154446023]	✓	✓	✓	✓	✓	✓
IGSF9B (intron)	Chr 11[133814743..133814771]	✓	✓	N.A	✓	✓	✓
TSPAN11 (exon)	Chr 12[31079849..31079877]	✓	✓	✓	✓	✓	✓

Figure 4.1: Potential off-targets of the *FGFR2* ZFN pair

A) Result from the ZFN site website (http://ccg.vital-it.ch/tagger/targetsearch.html). When a nucleotide mismatch is found at a given position between query and hit, the mismatched position is highlighted and underlined; the original nucleotide being displayed underneath (red).The spacer sequence size is represented by Ns (green). Results also show the number of mismatches between queries and mismatch site, and the genomic locus of the off-target. B) Sequencing results of the off-target ZFN binding site for each clone. The tick symbol means that the sequence was identical to database, proving that the ZFN did not cut that locus. N.A. means that some sequencing reaction failed to give a sequencing trace. See Appendix 8 for the entire sequencing data.

genome instability (Hanahan and Weinberg, 2000; Hanahan and Weinberg, 2011) along with an increased rate of mutations as compared to normal cells. It is therefore expected that the cell lines cannot stay true isogenic cell lines for an extended period of time and that, as passage number increases, the risk of accumulating additional point mutations increases as well.

4.1.1. Proliferation, cell cycle analysis and migration

In order to detect the impact of the single nucleotide change, the heterozygous clones (Het 1, Het 2 and Het 3) were compared to their control counterparts (Ctr 1, Ctr 2 and Ctr 3) in a series of *in vitro* assays.

The MCF7 cells are a weakly metastatic breast cancer cell line usually growing in clusters and not very motile, retaining contact inhibition. At confluence, they present a cobblestone morphology, typical of epithelial cell lines. The appearance of the sublines varied moderately between each other (Fig. 4.2A) but did not correlate with the rs2981578 genotypes. Het 1, Ctr 1 and Ctr 3 displayed more membrane elongations, resembling filopodia, observed at the edges of cell clusters, compared to the other clones. It has often been reported that MCF7 cells, like many cell lines (Wenger *et al*, 2004) have a tendency to deviate from their initial phenotypes as the number of passages in culture increases, and this discrepancy in phenotypic appearance may also have been caused by the stressful single cell cloning process.

The six clones were first subjected to cell cycle analysis using PI staining followed by flow cytometry. The heterozygous clones displayed a normal cell cycle profile, similar to the wild-type controls (Fig. 4.2B). Although their proliferation rate appeared reduced compared to MCF7 cells which had not been subject to transfection or single cell cloning (data not shown), directly after single cell cloning, MTS assay (Fig. 4.2C) and anti-Ki67 staining (Fig. 4.2D) did not reveal any significant differences between the clones themselves.

Additionally, migration was investigated using the organotypic model combined with a wound assay. The MCF7 cells do not normally invade during an organotypic

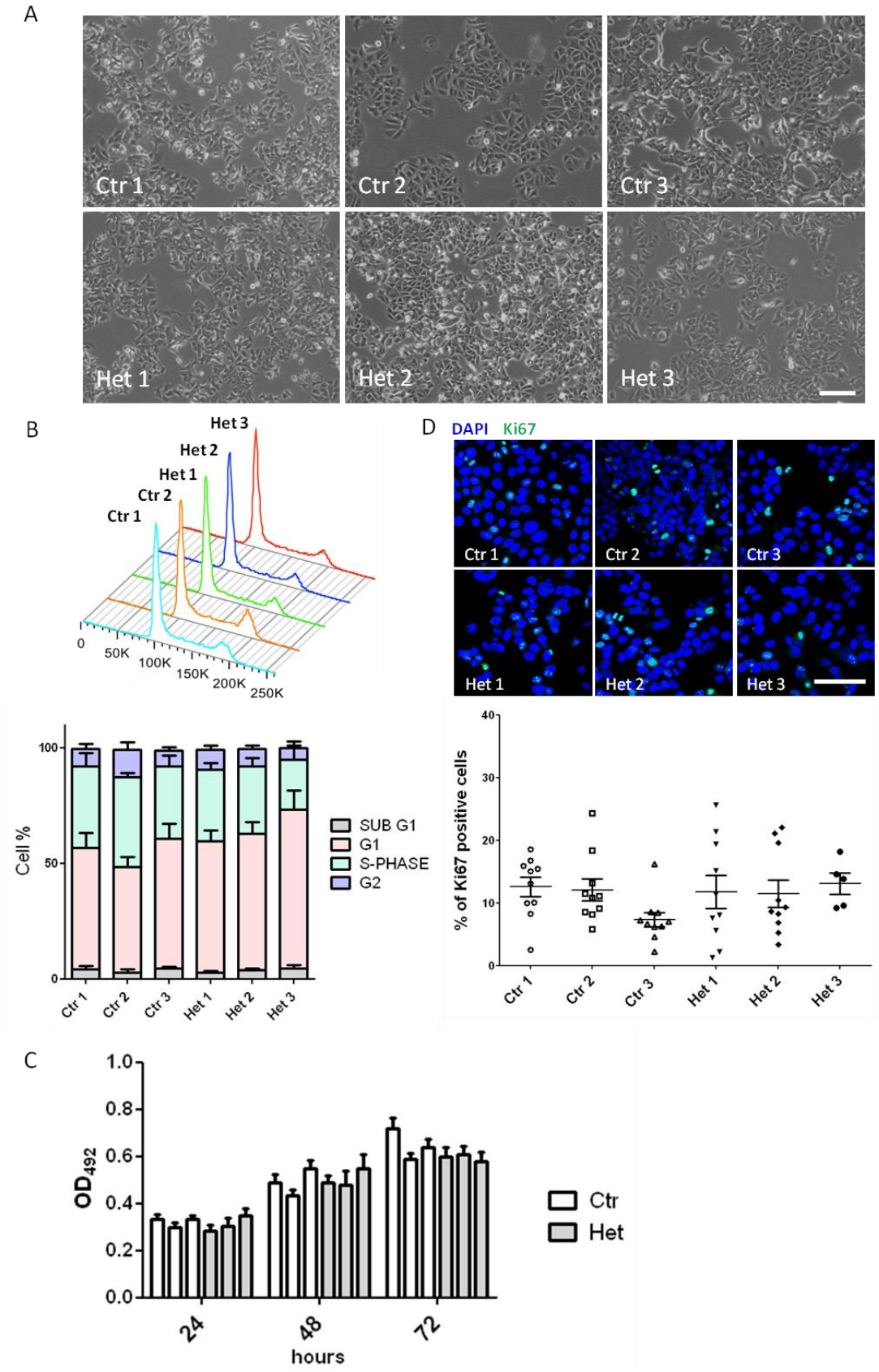


Figure 4.2: Characterisation of the heterozygous MCF7 clones, as compared to homozygous controls

A) MCF7-derived clones grown as a monolayer at passage 4 (x100). The ZFN-modified clones were derived from the MCF7 cell line, an adherent luminal epithelial cell line

isolated from pleural effusions from a woman with invasive breast carcinoma. The MCF7 cell line has a classic epithelial appearance in culture, usually growing as clusters retaining contact inhibition. Bar: 50 microns. B) Cell cycle analysis by PI staining and flow cytometry. DNA fluorescence signal showed a normal cell cycle DNA content profile for the ctr and the het clones. Two-way Anova statistical test was performed but did not reach significance with a p value of 0.1293. C) MTS assay comparing the cell number of the control clones and the heterozygous clones over 72 hours. Each bar represents an average of three independent experiments by indicating the mean of absorbance measured at 492 nm and the SEM. D) Proliferation was assessed using Ki67 staining on fixed cells. Quantification was performed by counting the percentage of positive cells in 10 fields (on average 976 cells/10 fields) of view for each clone. Mean ± SEM of three independent experiments are presented. Bar: 50 microns.

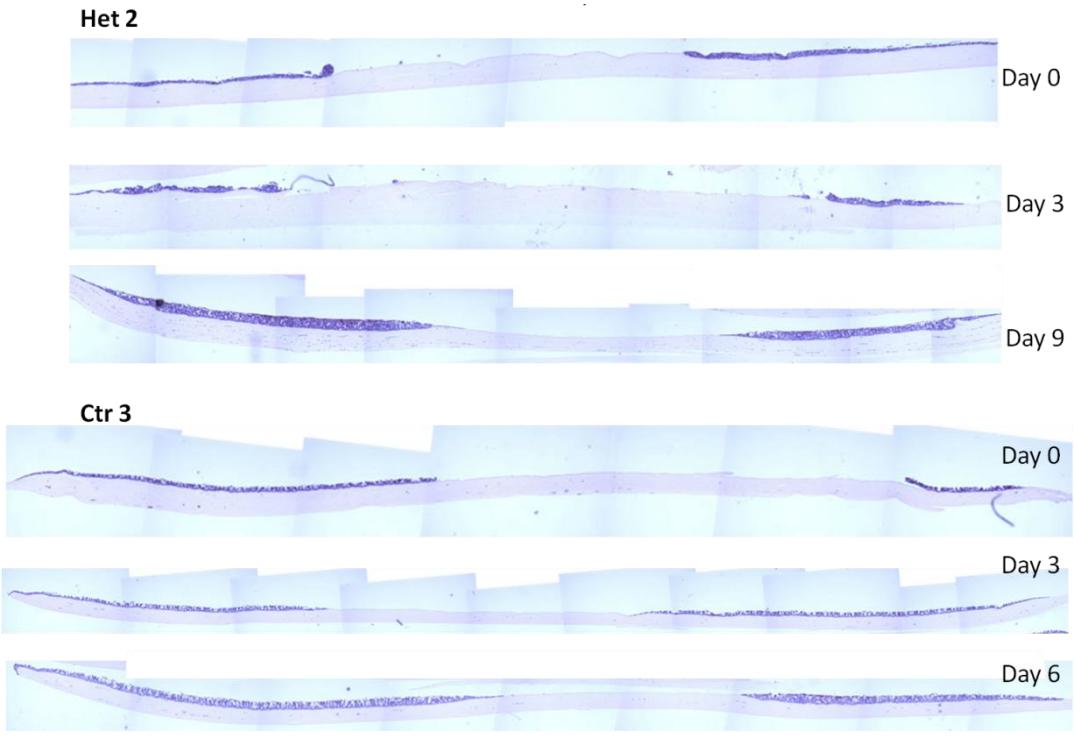


Figure 4.3: Migration assay using organotypic culture: Wound assay

The wound healing process was examined in Ctr 3 and Het 2 MCF7-derived clones after 0, 3, 6 and 9 days post wounding (with a punch biopsy). The MCF7 cells do not present any invasive phenotype in organotypic assay, and instead, proliferate on the top of the collagen/matrigel matrix. The cell layer sometimes detaches from the gel during the process of making slides. The different wound sizes at day 0 are therefore not comparable and no measurement of the wound closure can be undertaken.

assay (unpublished data), but instead, proliferate on top of the collagen/matrigel matrix and form, after a few days, a thick cell layer. This cell layer was wounded with a punch biopsy and the cells in the wound were removed (Fig. 2.2). The wound was then left to close for two weeks, allowing cells to migrate to the newly formed space. However, during the fixation of the organotypics and the embedding process at the end of the assay, thick cell layers occasionally detached (partially of entirely) from the matrix, making the measurement of the wound unrealiable. This assay was not best suited for the non-invading MCF7 cells, however, clear cell migration (cells moving as a single file at the edge of the wound) and wound closure could be seen in both Ctr 3 and Het 2 cells from day 6 (Fig. 4.3). The migrating rates could unfortunately not be compared between the controls and the heterozygous cells.

4.1.1. FGF and ERα signalling

FGF signalling and oestrogen receptor alpha expression were investigated. As demonstrated in Chapter 3, MCF7 cells expressed FGFR2, predominantly the epithelial-associated isofom FGFR2-b. Real-time PCR showed that no statistically significant difference was found, in terms of isoform levels, between the control and the heterozygous clones (Fig. 4.4A).

Two of the clones (Ctr 3 and Het 2) were chosen to be used in cell-based assays, on the basis of their most similar cell cycle profiles (Fig. 4.2B). The two cell lines were stimulated with 100 ng/ml of FGFR2-b specific ligands: FGF7 and FGF10, for increasing amounts of time and in the presence of heparin. The results showed that both FGF10 and FGF7 elicited robust ERK phosphorylation sustained after 60 minutes of stimulation (Fig. 4.4B) in both Ctr 3 and Het 2 cells. In a second type of experiment, the sensitivity of the receptors toward decreasing amounts of ligands was assessed. Only key time points of zero, 30 minutes and one hour were chosen. Although phospho-erk relative intensity was decreased, even the smallest amount of ligand (1 ng/ml) was capable of eliciting a signal and leading to ERK phosphorylation, demonstrating no apparent change in receptor affinity for the ligands (Fig. 4.4C). Interestingly, basal erk phosphorylation seemed reduced in Het 2 clones compared to the ctr 3 clones (Fig. 4.4B and C). In addition, the levels of

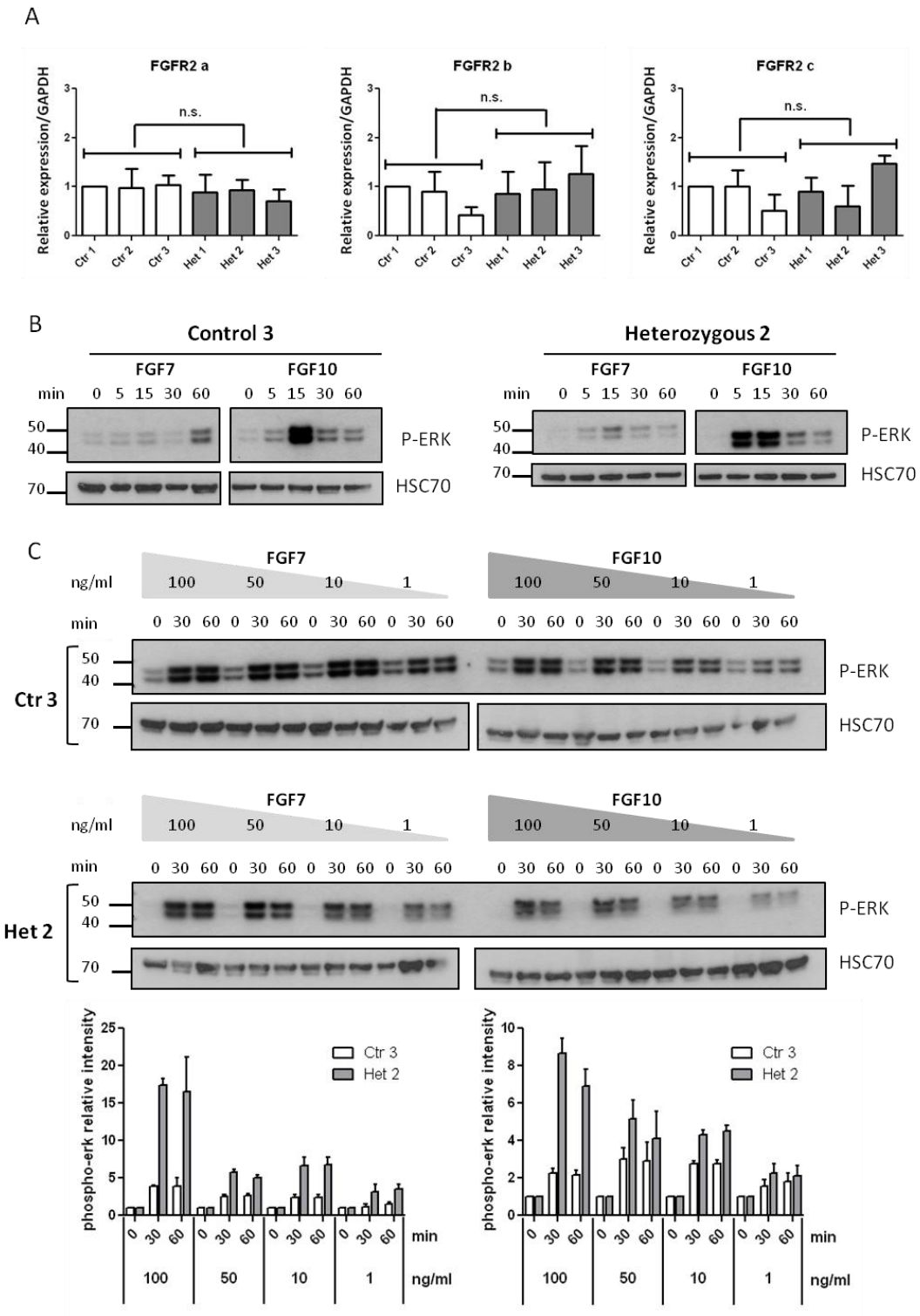


Figure 4.4: FGFR expression and signalling

A) The MCF7 clones were subjected to real-time PCR to assess the level of expression of the three different FGFR2 isoforms. No statistical difference was observed between the controls and the heterozygous clones (2-way ANOVA). B) The MCF7 clones were stimulated with two FGFR2-b specific ligands (100 ng/ml) from 5 min to 1 hour. Two different exposures are showed. C) Erk phosphorylation after stimulation of the Ctr 3 and Het 2 cells with decreasing concentration of FGF7 or FGF10 ligands. The experiments were repeated independently three times and a representative blot is shown, as well as the densitometry analysis for all of the experiments.

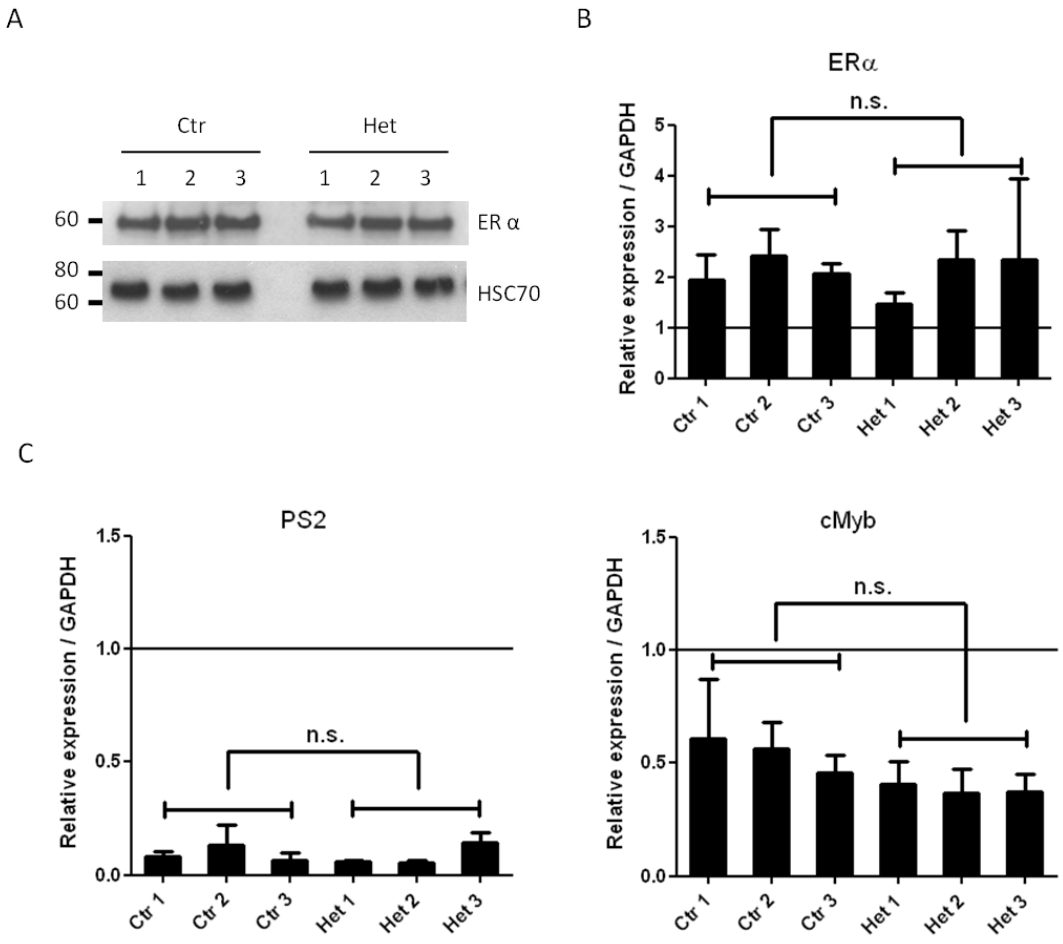


Figure 4.5: Oestrogen receptor alpha level in the MCF7 clones and response to Tamoxifen treatment

A) Western blot analysis of ERα expression level in the control versus the heterozygous clones. HSP70 was used as loading control. B) Quantitative RT-PCR of ERα level upon exposure to 1 µM Tamoxifen (TAM) relative to control (vehicle, EtOH) for 48h, and effect on the expression level of two ERα response genes: *PS2* and *cMyb*. mRNA levels are shown here relative to GAPDH expression, and normalised over control. Mean ± SEM of three independent experiments are presented. Statistical analysis (2-way ANOVA) showed that the Tamoxifen treatment significantly changed the mRNA level of ERα, PS2 and cMyb ($p<0.0001$) but no significance was reached when comparing the control clones and the Het clones (ERα: $p=0.6491$, PS2: $p=0.1098$, cMyb: $p=0.2304$).

oestrogen receptor alpha were investigated as ER positivity constitutes the only significant tumour characteristic that was associated with the *FGFR2* SNP haplotype.Western blot analysis did not show any significant differences between the oestrogen receptor alpha levels in the heterozygous versus the homozygous controls (Fig. 4.5A). The cells were treated with 1 μM Tamoxifen for 48 hours, which led to a significant increase in the level of ERα compared to the untreated controls. This increase has been reported in the literature as a response from a positive feedback loop (Carroll *et al*, 2003) but the exact mechanism remains unknown. Tamoxifen treatment, however, significantly reduced the level of ERα target genes such as *PS2* and *cMyb* ($p<0.0001$) (Gudas *et al*, 1995; Kim *et al*, 2000). This reduction was equivalent in both control and heterozygous clones, as shown by the lack of any statistically significant difference between the two groups, as assessed by 2-way ANOVA test.

From these first observations, the risk allele of rs2981578 does not seem to directly affect the expression level or the signalling of the *FGFR2* gene nor the level of oestrogen receptor alpha expression.

4.1.2. Transcription factor binding at the rs2981578 locus

a. Runx2 transcription factor

Runx2 was identified as the transcription factor mediating the increase in FGFR2 expression in cell lines with the disease associate allele of rs2981578 (Meyer *et al*, 2008). It was demonstrated in this study that exogenous Runx2 was able to bind the promoter of a *Luciferase* reporter gene on a site containing multiple repeats of the disease associated allele and its surrounding sequence. The disease associated allele at the Oct1/Runx2 site stimulated transcription 2 to 5 fold over the non-disease associated allele, independently of orientation. The ChIP data were less conclusive and only showed a relative increase in Runx2 binding from 0.8 to 1.4 (Meyer *et al*, 2008).

The attempts to replicate the Runx2 ChIP data for the rs2981578 locus in the MCF7 clones failed to show a significant enrichment of the binding of Runx2 to the rs2981578 locus (data not shown). Additionally, the ChIP experiements for Runx2

were carried out with a commercially available kit that used IgG controls only, for data normalisation. The apparent fold enrichment observed could have been an artefact caused by the use of IgG control for normalisation. Future experiments were carried out using internal positive and negative controls as well as the input DNA for normalisation of the real time PCR results.

b. Identification of other potential *trans*-acting factors

As Runx2 ChIP analysis failed to provide robust data, publicly available online ChIP-seq (chromatin immunoprecipitation followed by high-throughput sequencing) data on whole-genome scale (ENCODE, 2012) were used to identify other potential transcription factors present at the rs2981578 locus. Whole genome data from MCF7 and HepG2 (hepatocellular carcinoma) cell lines revealed that the transcription factor E2F-1, involved in cell cycle control (Muller *et al*, 2001) and the pioneer factor FOXA1 (Cirillo *et al*, 2002) were bound to the DNA at this locus.

E2F-1 is a transcription factor which acts predominantly as a regulator of the cell cycle by coordinating the expression of genes during early cell cycle progression (Takahashi *et al*, 2000). The E2F family works in tandem with the retinoblastoma tumour suppressor (*RB1*) to allow the entry of cells into the S phase of the cell cycle. It can also induce apoptosis in response to DNA damage. Because E2F-1 is such a key player in regulation of cell growth or cell death it has, not surprisingly, been implicated in human cancers such as lung (Park *et al*, 2012) and breast cancer (Xu *et al*, 2013).

The second transcription factor, *FOXA1*, that was identified in the ENCODE database, constituted an ideal candidate for studying the link between *FGFR2* intronic SNPs and increased risk of ER-positive breast cancer. Indeed, *FOXA1* is a pioneer factor responsible for opening the condensed chromatin for easy access by other transcription factors and has been shown to play an important role in maintaining euchromatic conditions and to be required for ERα binding (Carroll *et al*, 2005). The binding of FOXA1 to the rs2981578 SNP locus was confirmed in MCF7, T47D and ZR75-1 cell lines by ChIP-seq data analysis from a study on FOXA1 and oestrogen receptor function in breast cancer (Hurtado *et al*, 2010) (Appendix 6). Interestingly, Ross-Innes and colleagues (2012) have shown that ERα binding is

a dynamic process and that new ERα-binding sites were unique to seven patients with poor outcome as compared to eight patients with good outcome. When using the ChIP-seq data from that study, ERα was bound a few hundred base pairs away from the rs2981578 locus and only in samples associated with poor outcome (Appendix 6). The current hypothesis on the role of *FOXA1* in breast cancer is that *FOXA1* is capable of mediating a reprogramming of the ERα binding site (Ross-Innes *et al*, 2012). Considering the strong association with ER positive breast cancer risk and the minor allele haplotype of *FGFR2*, further ChIP assays were performed to assess the binding of FOXA1 in the ZFN-modified MCF7 cells.

As a pioneer factor, FOXA1 is capable of binding closed, condensed chromatin, which is transcriptionally inactive. The MCF7 clones (het 2 and ctr 3) were, therefore, either cultured in full medium or starved of oestrogen for 4 days and stimulated (or not) with 100nM of β-oestradiol (E2) for 1 hour, prior to chromatin isolation and ChIP analysis. A site within the fourth intron of CCND1 (Cyclin D1) and the Greb1 (growth regulation by oestrogen in breast cancer 1) promoter were used as negative and positive control, respectively, for FOXA1 binding (Ross-Innes *et al*, 2012). As expected, control cells showed enhanced binding of FOXA1 to the Greb1 promoter following oestrogen stimulation. Heterozygous cells showed relatively lower enrichment of FOXA1 binding. Despite an unexpected high level of FOXA1 binding to the Greb1 locus in heterozygous cells growing in full serum, the cells still showed a positive response of FOXA1 binding to the Greb1 promoter following oestrogen stimulation (Fig. 4.6A). Control cells showed significantly enhanced FOXA1 binding at rs2981578 relative to heterozygous clones in all culture conditions, but most achievably following ERα stimulation. Total FOXA1 levels were equal in both control and heterozygous cell lines (Fig. 4.6B).

a. Allele specific binding of FOXA1

In order to assess if the reduced binding of FOXA1 at the rs2981578 locus was caused by the rs2981578 risk allele, allele specific binding (ASB) was investigated using FOXA1 ChIP material in a specific SNP genotyping Taqman assay for rs2981578. The results indicated that all the cell lines tested showed the correct

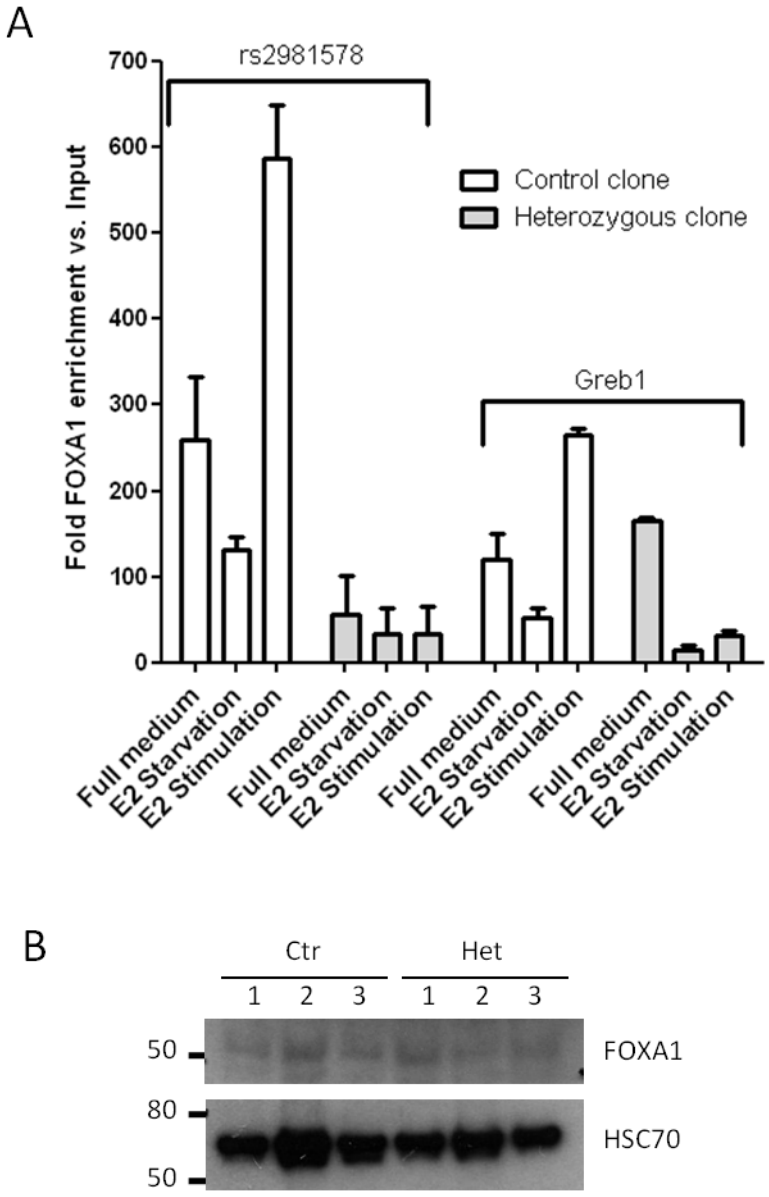


Figure 4.6: ChIP analysis of FOXA1 binding at rs2981578 locus in the SNP-edited MCF7 clones

A) FOXA1 ChIP-PCR results in heterozygous (grey) and control (white) clones, in full medium, upon β-oestradiol (E2) starvation or after 1h of E2 stimulation (100 nM). Greb 1 was a positive control FOXA1 binding locus (primers located in *Greb1* promoter region). The fold enrichment was normalised to a negative control (*CCND1* primers, located in an intron) and to the Input DNA for each sample. Error bars represent SEM of three independent experiments. B) Expression of FOXA1 in the controls and heterozygous clones. HSC70 was used as loading control (western blot is representative of three independent experiments).

segregation in the different regions of the graph (in accordance with their genotype), except for the MCF7 clones ChIP material, which all appeared to be heterozygous (Fig. 4.7). This artefact was caused by the combination of ChIP and Taqman assays, as the Taqman assay alone, using genomic DNA, showed correct genotyping information (Fig. 3.8 and Fig. 3.9). It was therefore impossible to distinguish the control clones from the heterozygous clones and therefore no assumption could be made regarding any ASB. The PCR products resulting from the Taqman/ChIP assay were cloned into cloning vectors and sequenced. However the results obtained did not reflect the preferential binding of a particular allele, and only constituted an equal repartition of each of the two alleles. This indicated that the Taqman probes were sometimes cloned into the vector and not the actual ChIP PCR product. This approach of assessing allele specific binding was therefore abandoned and another strategy using patient tissue samples, discussed in Chapter 5, was adopted.

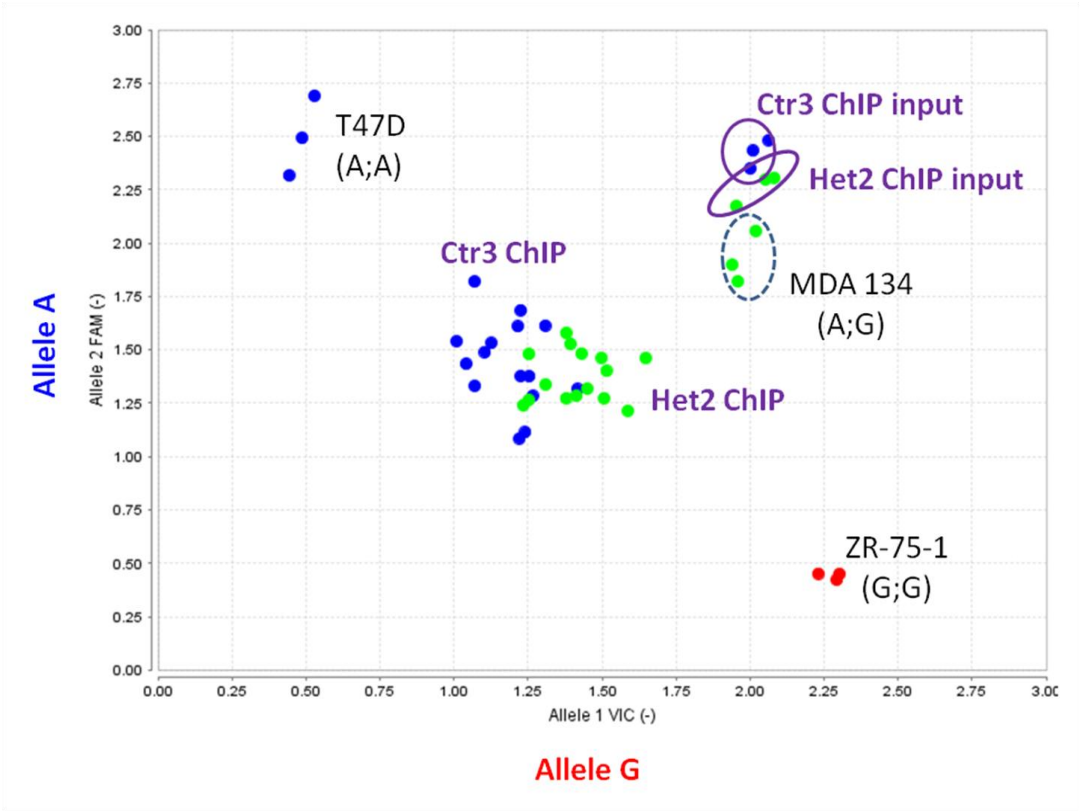


Figure 4.7: rs2981578 specific Taqman assay following FOXA1 ChIP

In order to assess allele specific binding of FOXA1, the input FOXA1 ChIP DNA samples from T47D, MDA-MB-134 and ZR-75-1 cell lines were used as controls for the three possible genotypes of rs2981578. ZF-75-1 (red), MDA-MB-134 (green) and T47D (blue) input samples segregated to the correct position of the graph (i.e. according to their respective amount of allele A and/or allele G). The ChIP samples of Ctr 3 and Het 2 MCF7 clones (purple circles), used in fig. 4.6 (E2 stimulation), were also used for Taqman assay in order to determine if FOXA1 displayed allele-specific binding. However, the het and ctr samples all grouped to the same region of the graph (behaving like heterozygous samples, even in the case of control samples), indicating a technical problem with this specific assay.

4.1. Discussion

Stable transgene expression is one of the most powerful genetic tools available when it is necessary to investigate the function of genes and the impact of genetic alterations such as oncogenic mutations. The generation of stable cell lines containing a randomly inserted transgene using recombinase-mediated cassette exchange or integrase-mediated site specific insertion (Sorrell and Kolb, 2005) is appropriate when studying a monogenic disease. However, expression levels between clones vary greatly due to chromosomal positions and copy number variations (Recillas-Targa, 2006), making the screening for identical clones laborious. In addition, random insertion may lead to genome alterations, such as gene inactivation, which may alter cellular phenotypes. The ability to use targeted, tailored changes, introduced into the genome using ZFN technology can overcome these limitations and can, potentially, allow the investigator to ask more precise biological questions.

The functional significance of the presence of *FGFR2* risk alleles is currently unknown. It was hypothesised that one SNP from the risk haplotype was functional and allowed the *de novo* binding of transcription factors that altered the FGFR2 expression level (Meyer *et al*, 2008). A specific FGFR2 ZFN pair was used to introduce the risk allele in MCF7 cells and the modified clones obtained were used in a series of *in vitro* assays and compared to unmodified MCF7 control clones.

The combination of zinc finger moieties used for target sequence specificity may not be completely specific; thus this might induce some cytotoxicity if off-target cleavages occur in essential areas of the genome. No off-target cleavage, and subsequent base pair deletions, was observed at seven possible off-target loci, in any of the controls or heterozygous MCF7-derived clones. The high specificity of the obligate heterodimers formed by the zinc fingers ZFNs, and their transient nature (mRNA rather than DNA) greatly reduces the possibility of off-target effects due to unspecific cleavage or stable integration to the genomic sequence. The spatial constraint of the 6 bp spacer sequence added an additional degree of specificity.

No marked or measurable changes (in cell cycle and proliferation markers) were detected between the two populations of MCF7 clones (Fig. 4.2B, C, D). Additionally, the lack of differences after FGF stimulation of Ctr 2 and Het 3 clones indicates that the receptor function and downstream signalling was not altered by the SNP modification, as expected given the intronic location of the SNP.

Unlike the results published in the Meyer study, no substantial Runx2 binding was observed at the rs2981578 locus. Additionally, an attempt to use a positive control for Runx2 binding, in the *β-casein* promoter also failed (data not shown). This might be due to the Chromatin Immunoprecipitation technique used for Runx2, using a commercially available kit rather than the in-house method developed by the Carroll lab (CRUK, CRI) and used in the FOXA1 ChIP experiment. For instance, the beads used for immunoprecipitation were not pre-incubated overnight with the antibody, but were added to the cell lysate and antibody at the same time. Also, the absence of a negative control for normalisation (like *CCND1* for FOXA ChIP) might explain the apparent fold increase in binding, but might only reflect the non-specific background binding of the antibody.

The pioneer transcription factor FOXA1 was therefore a much better candidate to explain the function of rs2981578. First, FOXA1 is required for ERα/chromatin interactions (Carroll *et al*, 2005) and plays a crucial role in reprogramming ERα target sites during cancer progression (Ross-Innes *et al*, 2012). Using data from the Ross-Innes study we were able to confirm that a FOXA1 binding site was present in the second intron of *FGFR2* in three breast cancer cell lines (ChIP seq data publicly available from Ross-Innes *et al*, 2012) (Appendix 6). FOXA1 ChIP showed a reduced binding of FOXA1 to the SNP locus in the heterozygous clone, whereas a very strong binding was observed in the control cell line (Fig. 4.6). FOXA1 is crucial in mediating the binding of ERα to its target genes, and whole genome ChIP-seq screening has demonstrated that FOXA1 plays a role in the reprogramming of ERα binding sites during breast cancer progression (Ross-Innes *et al*, 2012; Cowper-Sal lari *et al*, 2012). Interestingly, Ross-Innes and colleagues (2012) have shown that ERα binding is a dynamic process and that new ERα-binding sites were unique to seven patients with poor outcome as compared to eight patients with good outcome. When using the ChIP-seq data from that study, ERα was bound a few

hundred base pairs away from the rs2981578 locus and only in samples associated with poor outcome. The current hypothesis regarding the role of FOXA1 in breast cancer is that FOXA1 is capable of mediating a reprogramming of the ERα binding site (Ross-Innes *et al*, 2012). The role of each individual SNP forming the *FGFR2* haplotype, or their collective effect, on the dynamics of FOXA1 binding at the *FGFR2* locus remains to be elucidated. However, we could not confirm that the binding was allele specific for either the risk or non-risk allele as the SNP genotyping assay did not work on the MCF7 cell following ChIP (Fig 4.7).

The Meyer study (2008) hypothesised that increased FGFR2 expression, mediated by Runx2 binding at rs2981578 locus, was what underpinned the increased cancer risk. Howerver, no differencies in FGFR2 expression were detected between the SNP modified cell lines (Fig. 4.4A). Therefore, if *FGFR2* is indeed acting as an oncogene, one might expect no differences in cell behaviour between the clones. This suggests that rs2981578 risk allele alone is not capable of giving any advantage in cell growth and mediating the increase in breast cancer risk. A new hypothesis must be formulated to include other FGFR2 risk variants (from the same haplotype) and what appear to be the key players: *FGFR2*, *ERα* and *FOXA1*.

To this day, the only other ZFN-based approach used to correct two point mutations (with no additional footprint on the DNA), was carried out in patient-derived induced plutipotent stem cells to correct mutations in a gene associated with early onset Parkinson disease (Soldner *et al*, 2011). These non-synonymous mutations were localised to the coding region of α-synuclein (A53T and E46K), and the modified cells were shown to have kept their pluripotent characteristics but were not used for functional studies. The consequences of these point mutations in term of cellular behaviour were therefore unknown. However, it is possible to anticipate that disrupting the coding regions of a gene might have serious consequences on the mutated proteins themselves and their partners. Altering any *cis*-acting regulatory elements, might also have an impact on the cell behaviour but only in a phenotypically very subtle manner. A recent study provides a good example of an unexpected negative result using ZFN. The authors used ZFN to abrogate *MALAT1* expression (a non-coding RNA marker of lung cancer metastasis) in human tumour cells and found that this loss neither affected proliferation nor

cell cycle progression in human lung or liver cancer cells; furthermore it had no impact on nuclear architecture (Eissmann *et al*, 2012).

The attention has now shifted from a single SNP, rs2981578, to a set of risk SNPs forming a haplotype to try to identify what is mediating the breast cancer risk in the *FGFR2* second intron. Another SNP, rs35054928 (-/C), has emerged as new risk locus that might become important for elucidating this mechanism. The combination of different SNP genotypes was found more significantly associated with increased risk of breast cancer, compared to any individual SNP (personal communication, Dr Kerstin Meyer, CRUK Cambridge). Recently, a groundbreaking study by the Lupien group has shown that risk-associated SNPs of breast cancer are enriched for FOXA1 binding sites, which influence the function of this transcription factor (Cowper-Sal lari *et al*, 2012). Given more time, this project could now progress toward changing several SNPs at the same time and investigating the binding pattern of FOXA1 in these new cell line models.

CHAPTER 5

ALTERNATIVE METHODS TO STUDY SNP RS2981578

5. Alternative methods to study SNP rs2981578

5.1. Introduction

The genomes of related species show a remarkable amount of conservation; however, this is not reflected by the great array of phenotypic diversity observed. It originally was hypothesised that this diversity was due to variation in gene expression rather than structural changes in the genes (King and Wilson, 1975). For instance, a study looking at mRNA expression differences in the brain of three humans showed that there was a similar level of differences between them as there is between humans and chimpanzees (Enard *et al*, 2002). Understanding the link between genotype and phenotype remains a challenge in evolutionary biology but also in the study of risk for complex diseases, such as breast cancer. Indeed, like mutations that affect the protein sequence, changes in gene expression can promote cancer development. For instance, subtle downregulation of the wild-type tumour suppressor *PTEN* was enough to trigger cell proliferation and increased susceptibility to tumour development in a mouse model (Alimonti *et al*, 2010). The detection and recognition of those regulatory sequences that affect gene expression levels are the new challenges of molecular biology, as demonstrated by the recent ENCODE Consortium effort (Bernstein *et al*, 2012): although the understanding of how non-synonymous mutations in coding regions affect tertiary and quaternary protein structures has increased, the effect of polymorphisms and indels in non-coding DNA remains unclear. A major reason for this is that they can have an impact at different levels (i.e. transcriptional, post-transcriptional and epigenetic). The analysis of variation in gene expression is complicated by the potentially small differences associated with alterations in a single allele of a gene.

5.1.1. Cause of allele specific expression

The concentration of mRNA is controlled by two kinds of factors, *cis* and *trans*-acting factors, named as such for their localisation in relation to the gene whose expression they are regulating. *Trans*-regulation is mediated via diffusible factors such as a protein or ribonucleic acid (i.e. transcription factors, microRNAs) and constitutes the major regulatory system (Schadt *et al*, 2003; Cheung *et al*, 2010)

whereas regulatory elements acting in *cis* regulate gene expression by directly altering the local genomic sequence (i.e. a mutation in the promoter or intronic sequence that alters a transcription factor binding site or impacts on methylation status) (Wray *et al*, 2003). Interestingly, c*is* regulatory elements are generally binding sites for factors acting in *trans*. *Trans*-regulation has the potential to influence the expression of a multitude of genes with a single factor and influences both alleles of a gene indiscriminately, whereas *cis*-acting regulatory variation can lead to differential allele specific expression (ASE), in which the expression of one allele differs from another in a diploid cell. It was established that 10 to 22% of human genes are differentially regulated in such a fashion (Zhang *et al*, 2009). Evidence for the medical importance of *cis*-acting polymorphisms has been provided by the discovery of disease susceptibility loci that are not associated with protein coding regions or splice sites (ENCODE, 2012).

Cis-acting variation may explain 25% to 35% of inter-individual difference in gene expression and also influence mRNA processing, stability and isofom splicing (Pastinen and Hudson, 2004). *Cis*-acting elements may be located within enhancer and silencer elements that can be several tens or hundreds of kilo base pairs up or downstream from the transcribed sequence, or within the transcript itself in introns. Allelic imbalance can be caused by non-coding regulatory DNA polymorphisms but also by coding polymorphisms, in the case of nonsense-mediated mRNA decay of transcript harbouring early stop codons (Francastel *et al*, 1999). However, modulation of gene expression by epigenetic factors, including differential acetylation of histones or DNA methylation, parental imprinting (Ferguson-Smith *et al*, 2003) and random monoallelic expression (Ohlsson *et al*, 1998), may all be mistaken with regulatory polymorphisms.

5.1.2. Methods for measuring ASE

Allele specific expression of a given gene results in a differential concentration of cytoplasmic mRNA, and can be measured *in vitro* or *in vivo*.

In vitro approaches consist of the use of synthetic reporter constructs containing different portions of promoters (previously characterised, or known to contain candidate regulatory polymorphisms) to monitor the transcriptional activity of

several distinct reporter mRNA transcripts. However this strategy is only valid when employing a candidate approach and must take into account the fact that promoter regions are often poorly characterised and fail to represent the entire complexity of a given promoter in a given cell type, often omitting long range regulatory sequences. Additionally, these studies do not take into account the influence of *trans*-acting factors in the cell system used for transfection of the reporter construct (Volpi *et al*, 2000).

In vivo approaches are a direct way of studying the expression of alleles in their normal environment, including genomic and chromatin context. During transcription, hetero-nuclear RNAs are synthesised from the two copies of the DNA template, each bearing the differences caused by any heterozygous intronic polymorphisms. This form of pre-mRNA only exists briefly before being fully processed into mRNA, in which introns are spliced out, erasing any trace of the heterozygous nature of the transcript. The presence of an informative polymorphism (or marker SNP) for each individual transcript is therefore required in order to trace back its allelic origin. Another problem is caused by the fragile and unstable nature of single stranded mRNA, thus measurements are commonly performed using amplified cDNA from tissues or cell lines of interest. Differences in expression as low as 1.2 fold can be detected between samples (Pastinen and Hudson, 2004). The main advantage of *in vivo* approaches is that they can be scaled up to include the whole genome.

The method of polymerase loading assay (haploCHIP) can be used for studying the role of transcription in causing allele specific expression, and is based on isolating transcriptionaly active DNA fragments by immunoprecipitating the active RNA polymerase II enzyme (Charles Knight, 2005). Another approach combines whole genome screening using gene expression microarrays or RNA high throughput sequencing with genetic variation markers in order to identify new *cis*-acting polymorphisms that affect phenotypes, such sites of variation are named expression quantitative trait loci (eQTL). eQTL mapping studies have been applied in several model organisms and humans (Brem *et al*, 2002; Morley *et al*, 2004; Chesler *et al*, 2005).

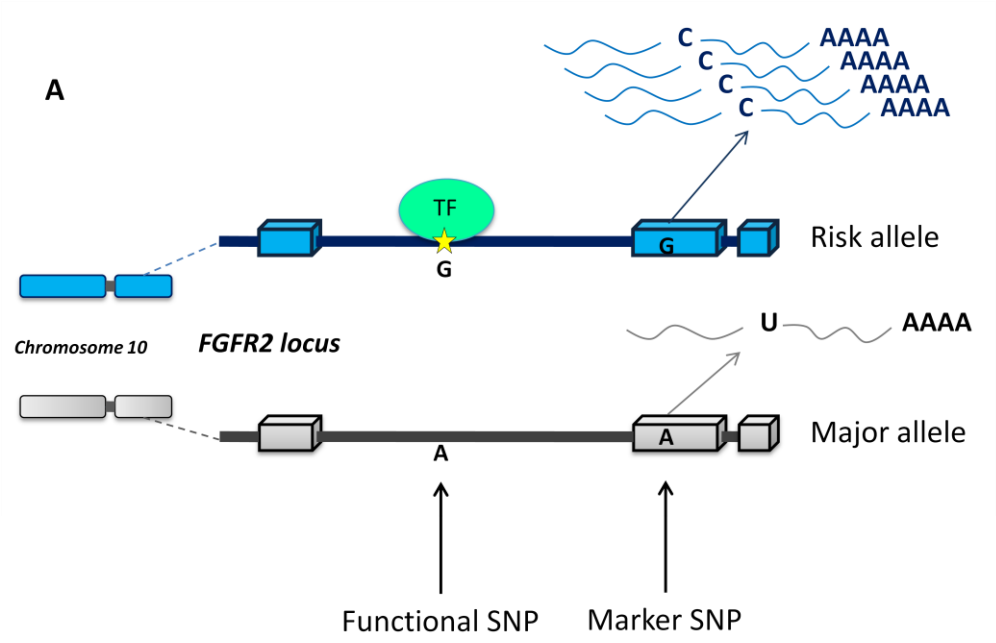
The focus of this study was to assess allele specific modulation of *FGFR2* expression and therefore study the *cis*-acting regulation of *FGFR2*. Data from the Meyer e*t al* (2008) study led these authors to hypothesise that the polymorphism rs2981578 creates a *de novo cis*-acting regulatory element in which binding of transcription factors on the risk allele led to increased *FGFR2* expression. In this chapter, the role of rs2981578 was investigated with an alternative approach to the ZFN-modified cell line approach. A quantitative genotyping assay was used to measure relative allelic abundance in breast cancer samples heterozygous for the rs2981578 SNP. Imbalanced allelic expression was determined by the heterozygous allele ratio of mRNA (cDNA) compared to the ratio of genomic DNA (1:1).

5.2. Results

5.2.1. Allele specific expression of *FGFR2*

Using the relative expression levels of variant SNP alleles within the coding region of a gene in the same sample (instead of using total mRNA levels originating from the two different copies of a gene) is an effective approach for identifying *cis*-acting regulatory SNPs (Milani *et al*, 2007). Since rs2981578 is intronic, and therefore spliced out of mature mRNA, the allelic origin of each mRNA molecule was tracked by looking at additional heterozygous SNPs in the coding region, named marker SNPs (Fig. 5.1A).

Potential marker SNPs located in the coding region of *FGFR2* were identified using the Ensembl Genome Browser website (Ensembl, 2010), by looking at the single nucleotide variants observed in the different FGFR2 transcripts. Among 327 total variations found in the coding sequence, 148 were synonymous variants and 179 were non-synonymous. Two of those variants were shortlisted, since they showed minor allele frequencies greater than 10%. The essential characteristic of a marker SNP is its heterozygosity, thus minor allele frequency is an important factor



B

Cell lines	rs2981578 (functional)	rs577593 (exon 5)	rs1047100 (exon 6)
AU561	-	A/A	C/C
BT 474	-	A/A	C/C
BT20	G/G	A/A	C/C
Cal51	-	A/A	C/C
H3396	A/A	A/A	C/C
MCF10A	G/G	A/A	C/C
MCF7	A/A	A/A	C/C
MDA-MB-361	-	A/A	C/C
MDA-MB-453	G/G	A/A	C/T
MDA-MB-468	A/A	A/A	C/C
SKBR3	A/G	A/A	C/C
T47D	A/A	A/A	C/C
ZR-75-1	G/G	A/A	C/C

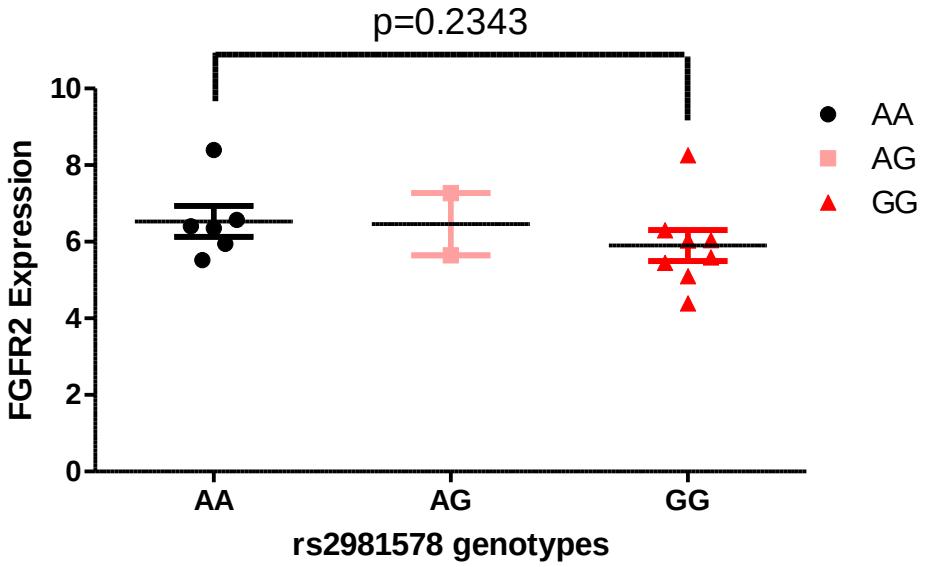
Figure 5.1: Allele specific binding (ASB) and Allele specific expression (ASE)

A) This diagram illustrates the concept of allele specific binding (ASB) of *cis*-regulatory elements in the context of a heterozygous functional SNP and consequent allele specific expression (ASE). The allelic origin of each mRNA molecule (blue or grey) can be traced by the use of additional heterozygous marker SNPs located in the coding region of the gene. TF stands for transcription factor. B) Genotypes for rs2981578 and two marker SNPs, rs577593 and rs1047100, in a panel of breast cancer cell lines.

because the greater the minor allele frequency, the better the chance of identifying heterozygous samples within cell lines or patient tissue samples.

rs1047100 is a synonymous SNP located in exon six of *FGFR2* (GT**A**/GT**G**). The nucleotide variance is at position Chr10:123298158 (GRCh37) and both variants encode for valine. The minor allele (A) frequency varies between 8% to 22% in the different populations of the 1000 Genomes project (1000Genomes, 2011) (Fig. 5.3). The second marker was the non-synonymous SNP rs755793 (A**T**G/A**C**G) in exon five, Chr10:123310871 (GRCh37). The ancestral codon, containing the thymine nucleotide, encodes for a methionine, which gets replaced by threonine, in the presence of the C allele. The minor allele (C) frequency varies greatly between populations, with a 36% frequency in African populations and an absence in European populations (Fig. 5.3). SNP rs1047100 was therefore used predominantly in this study to determine the allelic origin of the FGFR2 mRNA molecules, because of the more homogeneous allele frequencies across populations and the fact that this change does not affect the amino acid sequence of the protein synthesised from the mRNA transcript.

A panel of breast cancer cell lines was screened both for heterozygous functional SNP and marker SNP and analysed to detect allelic imbalance in *FGFR2* gene expression. These cell lines were all derived from patients of white European ethnicity, with the exception of MDA-MB-468, derived from a black women (Neve *et al*, 2006). The results showed that none of the 13 cell lines screened were suitable for the study as none were heterozygous for both functional and marker SNPs (Fig. 5.1B). The allele frequencies of rs2981578 and rs577593 observed in the cell lines were in accordance with the European population frequencies, where C: 53% and T: 47% for rs2981578 and G: 0% and A: 99.9% for rs577593 respectively. Concerning rs1047100, however, less than expected T alleles were found. The observed frequency was 3.8%, statistically significantly different from the population frequency of 22% (Binomial distribution, two tailed p=0.029), which indicate that obtaining such results from a panel of breast cell lines was highly unlikely (with a probability of 1.3%). The only cell line that showed heterozygosity for rs2981578 was SKBR3, an ERα negative cell line (Neve *et al*, 2006).



	Data source	
Cell line: TT	CCLE	Neve *et al*, 2006
MDA-MB-468	6.354	6.4021
MCF7	6.574	5.942
T47D	8.4	5.5236
H3396	N/A	N/A
Cell line: TC		
SKBR3	7.2668	5.6499
PMC42	N/A	N/A
Cell line: CC		
MDA-MB-453	5.100887	5.5976
BT20	6.0416	6.3042
MCF10A	N/A	8.2614
ZR-75-1	5.458313	6.0432
HCC70	4.393347	6.6932

Figure 5.2: FGFR2 expression levels in breast cancer cell lines according to their respective rs2981578 genotype (cell line based eQTL)

The rs2981578 SNP was genotyped in 11 breast cancer cell lines (green are ER positive and red are ER negative). The expression data (log_2 gene expression) were taken from Affymetrix expression microarray data publicly available using probe 2263_at (CCLE, 2012; Neve *et al*, 2006). SNP genotype data were then correlated with the expression level of FGFR2. A one-way ANOVA statistical test was carried out.

Additionaly, publicly available copy number variation (CNV) data were used to assess the level of *FGFR2* expression in different breast cancer cell lines, according to their rs2981578 genotype (Fig. 5.2). The result suggested that *FGFR2* expression was similar in all the cell lines for which CNV data was available. It is important to note that only 5 of the 11 cell lines were ERα positive (highlighted in green) and this approach could be informative if more ERα positive cell lines were included, as the rs2981578 locus appears to have no impact in an ERα negative setting.

Given the established limitation of using cell lines (too few in numbers and not carrying the adequate genotypes) (Fig. 5.1 and 5.2), tissue from patients with ERα positive breast cancer was interrogated. Breast tissue samples were obtained from the Breast Tissue Bank at Barts in collaboration with Prof Louise Jones (ethics approved ref no. 05/Q0403/199) and selected purely on the basis of ERα positivity, regardless of treatment and ethnicity (Table 6.1). Total DNA and RNA from 72 ERα positive breast tumours and their surrounding tissues were used and each sample was genotyped for rs2981578 and the two marker SNPs (Table 5.1). The allele frequencies of rs2981578 and rs1047100 in the patients' samples were representative of the overall population data from the 1000 Genomes project (Fig. 5.3). Allele G of rs755793 was represented at a frequency higher than predicted from population data, indicating a potential bias towards an increased number of patients with African descent in the sample set. However, only 8.3% of the patients were of a Black background compared to 68% of a White background (Table 6.1). Additionally, patients qualified as Asian in the sample set (composed of Indians, Bangladeshi and Pakistani patients) represented 10% of the samples and are not representative of the East Asian population (ASN) of the Hap Map or the 1000 genomes data bases, composed mostly of Chinese, Japanese and Vietnamese individuals. Little information is available as yet on SNP allele frequencies in Indian, Bangladeshi and Pakistani populations (SAN, south Asian super population code).

Five samples (15, 36, 39, 59 and 62 in red), which were heterozygous for both functional and marker SNPs, were selected for ASE analysis, and 10 (green) were used as controls (homozygous for rs2981578 and heterozygous for rs1047100) (Table 5.1).

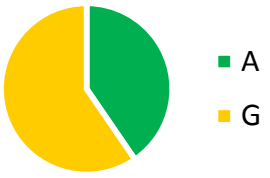
	Genotype		
Sample	rs2981578	rs1047100	rs755793
1	GG	CC	AG
2	AA	CC	AG
3	GG	CC	AG
4	GG	?	AG
5	AG	CC	AA
6	GG	CT	AG
7	GG	CC	AG
8	GG	CC	AG
9	AA	CC	AG
10	AA	CC	AG
11	GG	CC	AG
12	GG	CT	AG
13	AA	TT	AG
14	GG	CC	AG
15	AG	CT	AA
16	?	CC	AG
17	GG	CC	AG
18	AG	CC	AA
19	GG	CC	AG
20	GG	?	AG
21	GG	CC	AG
22	AA	CC	AG
23	GG	CC	AG
24	AG	CC	AA
25	AG	TT	AA
26	GG	CT	AG
27	GG	CT	AG
28	AA	CT	AG
29	GG	CT	AG
30	AG	CC	AG
31	AA	CC	AG
32	AG	CC	AG
33	AG	CC	AA
34	GG	CC	AG
35	AG	CC	AA
36	AG	CT	AA

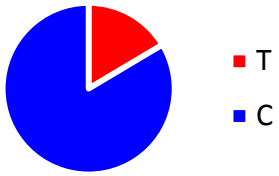
	Genotype		
Sample	rs2981578	rs1047100	rs755793
37	GG	CC	AG
38	AA	CT	AG
39	AG	CT	AA
40	GG	CC	AG
41	AA	CC	AG
42	AG	CC	AA
43	AA	CC	AG
44	GG	CT	AG
45	AA	CC	AG
46	?	CC	AG
47	?	CC	AG
48	AA	CC	AG
49	AA	CC	AA
50	GG	CC	AG
51	AG	CC	AA
52	GG	CC	AA
53	AA	CT	AA
54	AG	CC	AA
55	AG	CC	AA
56	AG	CC	AA
57	AG	CC	AA
58	AA	CC	AA
59	AG	CT	AA
60	GG	CC	AG
61	GG	CT	AA
62	AG	CT	AA
63	GG	CC	AA
64	GG	CC	GG
65	GG	CT	GG
66	AG	CC	AA
67	AA	TT	AA
68	GG	CC	AA
69	AG	CC	AA
70	AA	CT	AA
71	GG	CC	AG
72	AG	CC	AA

Table 5.1: Genotypes of a panel of breast cancer tissues

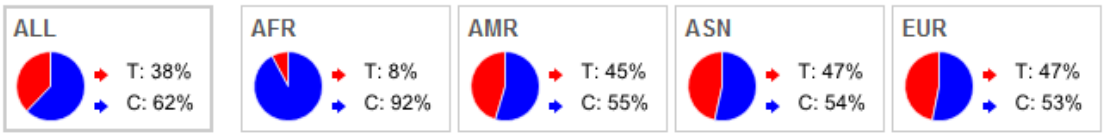
A panel of ER positive breast cancer tissues was genotyped, for both marker SNPs and rs2981578. Orange shaded positions indicate heterozygous samples. Grey shaded positions indicate ambiguous results, where the SNP status could not be determined. In red are the samples used for analysis and in green the ones used as controls.

rs2981578		rs1047100		rs755793	
A (T)	0.382	**T**	0.163	**A**	0.687
G (C)	0.612	**C**	0.812	**G**	0.319

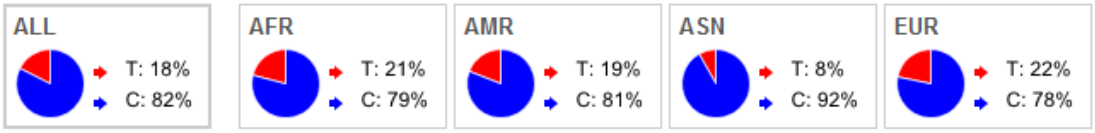




rs2981578 (1000 Genomes)



rs1047100 (1000 Genomes)



rs755793 (1000 Genomes)

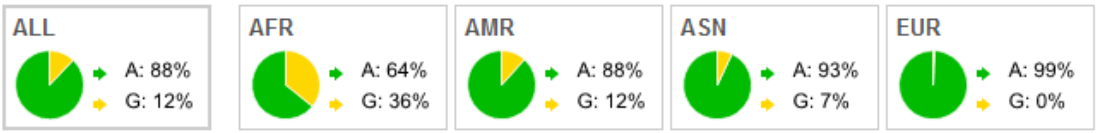


Figure 5.3: Allelic frequency in Barts Breast Tissue Bank samples

Allele frequencies measured in 72 ER positive breast cancer samples, compared to 1000 Genome data set. The three letters codes are Super population codes that regroup data from several populations: AFR: African, AMR: Ad Mixed American, ASN: East Asian, EUR: European. When the code ALL is used this means that all individuals from that the data set are being considered.

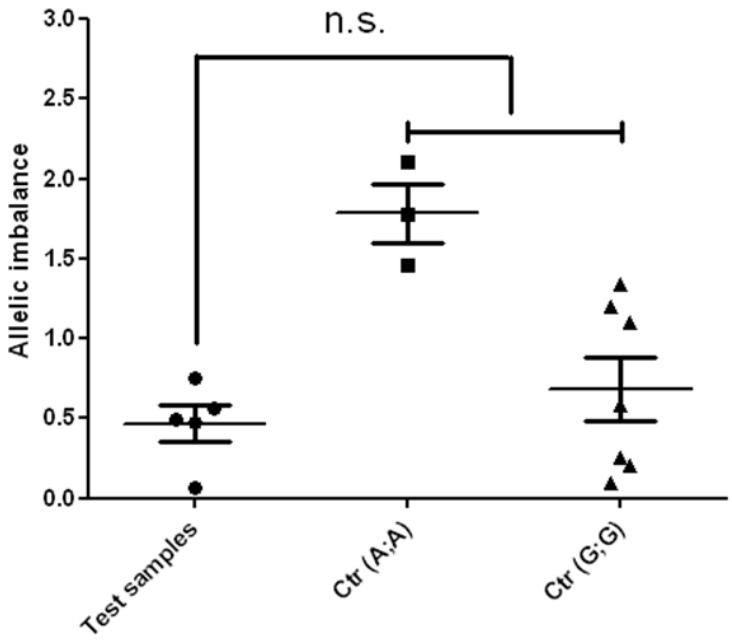


Figure 5.4: *FGFR2* allelic imbalance in breast cancer samples

Absolute differences between cDNA and gDNA Ct values as measured by SNP genotyping Taqman assay for SNP rs1047100. Samples were assayed in triplicate. Mann Whitney test was performed using Prism software, error bars represent SEM. Five test samples were analysed and compared to three homozygous controls with the major allele (A;A) and seven homozygous controls with the minor, or risk allele (G;G).

Real time PCR using allele specific Taqman probes was performed for each sample, using genomic (gDNA) and complementary DNA (cDNA) templates. Imbalanced allelic expression is detected when the heterozygous allele ratio in mRNA (cDNA) differs from the corresponding 1:1 ratio in genomic DNA. Cycle threshold (Ct) values obtained for both alleles of rs1047100 in cDNA were normalised to Ct values obtained from gDNA to correct for errors coming from potential copy number variations. The absolute differences between Ct values from cDNA and gDNA, were calculated (Fig. 5.4) (Appendix 9). Mann Whitney test (p=0.1645) indicated that the results did not show any significant difference in absolute levels of expression (i.e. allelic imbalance) in the heterozygous samples compared to controls (A;A and G;G genotypes). Surprisingly, the controls homozygous for the non-disease associated allele of rs2981578 were the ones that displayed the greater difference between expression of each allele.

5.2.3. Selection pressure: polyclonal population expansion

FGFR2 has been reported to act as an oncogene in breast cancer and increased FGF signalling might promote cancer initiation or progression by protecting the cells from apoptosis (Hishikawa *et al*, 2004) and stimulating growth and proliferation (Turner *et al*, 2010).

In order to test this hypothetical advantage, three heterogeneous populations (ZFN1, ZFN2, ZFN3) composed of a mixture of wild-type MCF7 (A;A) and ZFN-modified cells (A;G or G;G) were cultured over a period of 20 passages. The relative amount of each rs2981578 allele was measured over time using allele specific Taqman probes to monitor any changes in the proportion of the two different genotypes.

The Ct values revealed, as expected after ZFN genome editing, a predominant proportion of wild-type cells (with Ct values around 30 cycles), with a slight increase (2 cycles difference) in G alleles post ZFN transfection, that persisted for 3 passages (Fig. 5.5). However, the Ct values returned to the level of the control, untransfected cells rapidly and no additional changes in Ct values were observed. The apparent increase in G allele frequency at passage 17 was an artefact caused

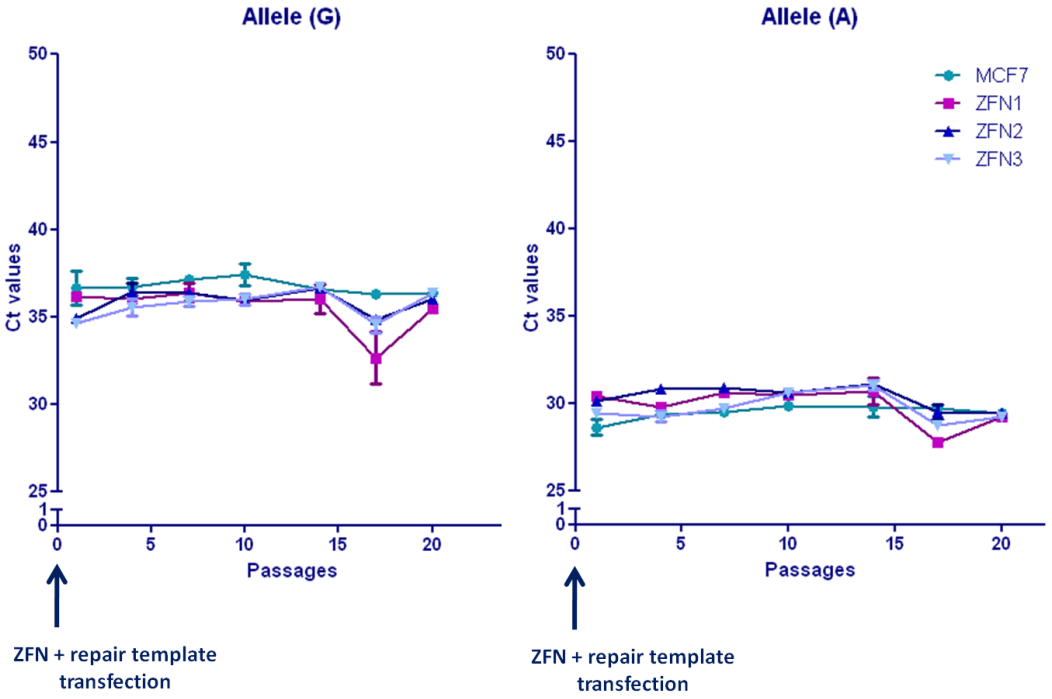


Figure 5.5: Ct values of each allele of rs2981578 (A or G) over 20 passages of ZFN-edited MCF7 cells

Three independent cultures (ZFN1, ZFN2 and ZFN3) of MCF7 cells were transfected with ZFN mRNA and MCF7 repair template (containing the risk allele) and kept in culture over a period of 20 passages. Genomic DNA was extracted every three passages and used for SNP genotyping Taqman assay, to monitor the levels of each allele of rs2981572. Results are represented as Ct values for each allele over time. Untransfected MCF7 cells were used as control.

by the poor quality of the genomic DNA samples, as this drop in Ct values was observed for both G and A alleles. The conclusion from this experiment indicates that the presence of the G allele in the *FGFR2* haplotype does not give a measurable growth advantage to the modified MCF7 cells in 2D culture.

5.3. Discussion

Two independent GWAS studies have identified *FGFR2* as a risk factor for breast cancer (Easton *et al*, 2007; Hunter *et al*, 2007). It was established that the disease associated alleles were inherited as a haplotype of eight SNPs in European populations, in which rs2981578 might constitute the causal variant (Meyer *et al*, 2008). In this chapter, the effect of the risk allele of rs2981578 on *FGFR2* expression was investigated. Measuring the relative expression of mRNA molecules originating from different copies of the same gene is a useful method to determine allelic imbalance caused by a *cis*-acting heterozygous polymorphism (Milani *et al*, 2007). The hypothesis under test was that the single nucleotide polymorphism, rs2981578, considered to be functional by allowing the *de novo* binding of transcription factors on the DNA molecule carrying the disease associated allele of the SNP, would regulate *FGFR2* expression.

Panels of ERα positive breast cancer samples and cell lines were genotyped to identify heterozygous SNPs (rs2981578, rs1047100 and rs755793). None of the breast cancer cell lines harboured the correct combination of heterozygous SNPs and were thus excluded from the study (Fig. 5.1B). Overall the allele frequencies measured in the breast cancer samples (Fig. 5.3) were very similar to those of the AMR (Ad mixed American), ASN (East Asian) or EUR (European) populations of the 1000 Genomes project, except for SNP rs755793, in which the G allele was represented at a high level in the samples (0.319) but was completely absent in the EUR populations. The limited information available on SNP allele frequencies of south Asian population (Indians, Bangladeshi and Pakistani individuals), constituting 10% of the data set in this study, might partly explain the discrepancy in the data. In the future, rs755793 might become useful as a marker SNP in similar studies of allelic imbalance in individuals of African descent, as its minor allele frequency reaches 36% in African population data. African descent population

(American or British alike) are a very singular group to study in relation to breast cancer risk as it was established that breast cancer presents itself at a younger age with distinct disease characteristics in this population (Bowen *et al*, 2008).

The allelic imbalance data showed that no significant difference in allelic expression levels could be observed in patient samples. This indicates that variation in allelic expression, if present, may manifest itself in a cell type or state specific manner, or that environmental conditions and/or physiological feedback mechanisms may mask the impact of subtle *cis*-acting variants on expression levels. An important consideration is the makeup of the tissue samples used for RNA and DNA extraction. The genetic material used for this experiment was actually a mixture of epithelial cancer cells, surrounded by stroma and blood vessels. One can therefore hypothesise that if the allele specific binding affects transcript production and consequent expression is cell type specific, any small effect would be diluted down by the presence of other cell types. Indeed, a study by Huijts and colleagues (2011) showed that FGFR2 mRNA levels were only increased in primary fibroblasts, but not primary epithelial cells from 98 breast cancer patients with rs2981578 risk allele (Huijts *et al*, 2011). Although the primary fibroblasts used were isolated from the skin of the patients and not the breast or tumour site, this indicates that the risk is mediated through a stroma-specific phenotype. Interestingly, it was demonstrated by Yan and colleagues (2002) that a smaller than anticipated number of normal individuals display ASE. They examined a normal population of 96 individuals, and interrogated heterozygous SNPs in 13 genes (Yan *et al*, 2002). Among all the heterozygous individual tested (from 17 to 37 depending on the genes), only a small minority showed ASE (3% to 30%). They also observed, by studying their pedigree, that altered allelic expression was an heritable trait. Additional patient samples that carry heterozygous functional and marker SNPs are therefore required to increase the power of this experiment as well as laser micro-dissection of tumour samples to obtain nucleic acids originating exclusively from a single cell population.

Observing an increase in the level of expression of an oncogene, here *FGFR2*, has little importance in terms of elucidating its impact on cancer incidence or progression without a concomitant phenotypic advantage for the cancer cells that

show such an increase. This advantage could be an increase in cell proliferation, as FGFR2 pathway stimulation is known to increase cell proliferation and migration in cancer cells by signalling via the MAPK signalling pathway (Ropiquet *et al*, 1999). However, no increase in the number of the cells carrying the risk allele of *FGFR2* SNP was observed in a population of cells composed of non-risk allele and risk allele genotypes, indicating that the risk is not mediated through an advantage in cellular proliferation, at least not in the context of cancer cells growing in 2D cultures without any other cell type present. Future refinements to this line of investigation could be to test behaviour in more physio-mimetic 3D culture models (Chioni and Grose, 2009) or testing response to cellular stress or insult, for example following serum starvation or chemotherapeutic challenge.

Identifying a true regulatory variant is complicated when a SNP shows very high linkage disequilibrium (LD) with other variants in close proximity. It is also possible that changes in chromatin structure, by epigenetic mechanisms, might also be an important factor to consider (Tirosh *et al*, 2008). It was shown by Zhu and colleagues that breast cell lines, harbouring the disease associated allele displayed histone acetylation at three SNP loci including rs2981578, hypothesising that this chromatin modification would modulate access to transcription factor binding sites and splicing sites (Zhu *et al*, 2009).

In conclusion, the phenotypic effect of rs2981578 remains unclear and might involve a cell type specific *FGFR2* regulation that could not be measured in tissue samples with mixed cell populations.

CHAPTER 6

GENERAL DISCUSSION

6. General discussion

6.1. Introduction

The development of breast cancers that are not the result of mutations in high penetrance susceptibility genes like *BRCA1* and *BRCA2*, are caused by a multitude of genetic factors, each conferring a small increase in the overall risk. This was demonstrated by multiple genome wide association studies and the *FGFR2* second intron was one of the most significant loci identified in ER positive breast cancers (Easton *et al*, 2007; Hunter *et al*, 2007). This locus was characterised by a haplotype composed of several SNPs in linkage disequilibrium within the large second intron of the gene. An early functional study hypothesised that rs2981578 was the functional SNP and that the risk was mediated via allele specific expression of *FGFR2*, as the result of differential binding of a trans-acting enhancer (Meyer *et al*, 2008). This enhancer was identified as the Runx2/Oct1 complex that has previously been shown to act in a similar fashion in *β-casein*, a mammary gland specific gene. The overall aim of this work was to create a set of isogenic breast epithelial cell lines to study the role played by rs2981578 in mediating breast cancer risk. To this end, ZFN technology was used as a means of editing rs2981578 in breast cancer cells. The second objective was to characterise the cell lines created and study the impact of rs2981578 *in vitro* in 2D cell culture models, thus deciphering the role played by *trans*-acting factors in mediating the breast cancer risk. Finally an alternative approach to study this mechanism was adopted, by using breast tissues from cancer patients with ER positive breast malignancy to study the allele specific expression of *FGFR2 in vivo*.

6.2. Creation of ZFN-edited breast cancer cell lines

Conventional methods for the study of gene function can be challenging and often use indirect approaches; for instance overexpressing the gene in a mammalian expression vector or knocking it down transiently using siRNA (Ahmed *et al*, 2010). For the study of *cis*-regulatory sequences, similar indirect methods are commonly used, such as reporter assays. ZFN-mediated genome editing presents several advantages over these conventional methods for generating modifications at the endogenous genomic DNA sequence, but can prove challenging.

The selection of a suitable candidate cell line for ZFN-mediated genome editing was a crucial step with potential repercussions during the entire editing process. The biological variability associated with the use of different human cancer cell lines might in part explain the differences in transfection efficiency, targeted DNA repair and also response to single cell cloning, all encountered in the genome editing process.

Additionally, the choice of potential ZFN binding sites was restricted to the immediate vicinity of the target SNP, which meant that the optimal ZFN pair was less efficient than they could have been, had the whole *FGFR2* locus been available for targeting. This resulted in a low efficiency of genome editing in the breast cancer cell lines tested. The initial optimisation process was therefore one of the most crucial steps in obtaining the model cell lines, and this, inevitably, took a considerable period of time. Several methods for transfection were tested, with Amaxa nucleofection being found to be the most suitable technique in MCF7 cells, further optimised by the introduction of a GFP control plasmid, co-transfected with the ZFNs and used for FACS enrichment of the GFP positive cells (Soldner *et al*, 2011). However, the 'cold shock' technique, that reportedly increased ZFNs cutting efficiency (Doyon *et al*, 2010), was not successful with *FGFR2* ZFNs. The problem of relative low efficiency of gene editing is common to many other studies and a lot of efforts are now being put into improving ZFN technology, as exemplified by recent reports suggesting the use of the proteasome inhibitor MG132 during the editing process as a way to increase the half-life of ZFN proteins (Ramakrishna *et al*, 2013), or the use of surrogate reporters that express GFP only when the reporter has been cleaved by the ZFN and a consequent frame shift mutation has occurred (Kim *et al*, 2011).

The choice of candidate cell lines was important not only for genome editing purposes, but also to provide the adequate phenotypic context in which to study the *FGFR2* SNP. As mentioned previously, ERα positivity was the single most important factor associated with the breast cancer risk mediated by *FGFR2* SNP, making the use of an ER positive cell line essential. Additionally, the increase in risk conferred by the *FGFR2* SNP was relatively small (1.63 for homozygous risk allele carriers) and was hypothesised to be easily masked by other major genetic

alterations commonly found in classic breast cancer cell lines. The immortalised but otherwise normal MCF10A cells (Soule *et al*, 1990) appeared to be the ideal cell line, but unfortunately lacked ERα expression. However, a study by Zhao and colleagues suggested that the MCF10A cell line had the potential to express ERα protein in an inducible fashion (Zhao *et al*, 2008). The results obtained in the Zhao study are in contradiction to the results presented in Chapter 3. It was demonstrated that the lack of ERα expression in the MCF10A cell line was not due to the presence of miR221 and miR222, as this cell line (and other derived cell lines) lacked ERα mRNA. MCF7 cells, which express high levels of ERα, were therefore selected as an alternative cell line for use in the study.

Three clones carrying one copy of the rs2981578 risk allele were obtained (none had a biallelic change after the first round of ZFN editing) and three other non-modified clones were selected as controls. A final attempt to edit the two copies of *FGFR2*, to obtain the final set of cell lines, using a different repair template (ssODN), failed. The potential off-target effect of the FGFR2 ZFN was evaluated and considered non significant as sequencing of seven potential off-target binding sites failed to show deletions due to NHEJ.

Cell-based assays showed that there was no change in cell cycle progression, nor any apparent advantage in cell growth or migration/invasion in cells carrying the risk allele of rs2981578 (heterozygous versus non-modified controls). Crucially, it was established that Runx2 was not the key transcription factor that mediated the rs2981578 risk, but instead, it was the pioneer factor FOXA1 that appeared more important in these studies. The FOXA1 chromatin immunoprecipitation experiment showed that there was a reduced binding of FOXA1 to the SNP locus in two out of three of the heterozygous clones, whereas a very strong binding was observed in two out of three control cell lines. It has now been established that FOXA1 is crucial in mediating the binding of ERα to its target genes, and whole genome ChIP-seq screening has demonstrated that FOXA1 plays a role in the reprogramming of ERα binding sites during breast cancer progression (Ross-Innes *et al*, 2012) (Cowper-Sal lari *et al*, 2012). The role of each individual SNP forming the *FGFR2* haplotype, or their collaborative effect, remains to be elucidated.

Finally, the assay that could have shed more light on the potential differential allelic binding of FOXA1 was the rs2981578 Taqman assay. The identification of which rs2981578 allele in an heterozygous sample that was pulled down predominantly after chromatin immunoprecipitation would have been very informative, but failed due to technical problems with the assay (discussed in Chapter 4).

6.1. Study of allele specific expression in a cohort of patient tumour samples

The cohort of patient samples collected at Barts Hospital, and examined in Chapter 5, did not show any allelic imbalance in FGFR2 expression. However, as pointed out in the chapter, the heterogeneous nature of the tumour samples used might explain the lack of allelic imbalance if that phenomenon is cell type specific. The examination of the patient genotype, and the fact that ASE is not present in every heterozygous individual, emphasised the importance of increasing the cohort of patients studied (Yan *et al*, 2002). Additionally, the composition of our patient cohort has revealed that genetic data on population originating from central and western Asia, such as India, Bangladesh and Pakistan, are currently missing from the main publicly available data bases such as the 1000 Genomes project, and that little information is available in term of SNP allele frequencies for these populations (Table 6.1).

6.2. Future work

6.2.1. *FGFR2* haplotype study

It has now become apparent that the breast cancer risk, in the context of *FGFR2* mediated risk, is attributable to a group of SNPs, called risk haplotype, rather than a single one (personal communication, Dr Kerstin Meyer, CRUK-CRI). Indeed, an unpublished genetic study conducted in Prof Bruce Ponder's lab has validated the importance of rs2981578 but only in association with other SNPs alleles. Further work on this risk haplotype could be made by the further modification of other SNPs such as rs35054928, 123 bp away from rs2981578. This SNP is in fact an insertion of a C risk allele (-/C), in which C is present 83% of the time in populations of European descent.

			%
Black	Black British	3	
	Caribbean	1	0.083333
	Other Black	2	
Asian	Pakistani	2	
	Bangladeshi	2	0.097222
	Indian	2	
	Other Asian	1	
White	Eng/Scot/Welsh	39	
	Irish	2	0.680556
	Greek	3	
	Other White	5	
	unknown	10	0.138889
	total	**72**	1

Table 6.1: Ethnicity of breast cancer samples

Proportion of each ethnicity within the 72 breast cancer samples obtained from the Breast Tissue Bank.

The risk allele is, however, not carried by the MCF7 cell line (Appendix 14). ZFNs could again be used in our modified MCF7 clones, along with ssODN repair templates to create an array of model cell lines with all the possible SNP combinations from that risk haplotype. The mechanism underlying the risk might be more complex than anticipated, and affect other genetic loci. For instance, a new study showed that SNPs at 16q12 (near the *TOX3* gene), associated with breast cancer risk, were located in enhancer regions and altered the binding affinity for FOXA1. This FOXA1 binding site on the chromatin was able to form a binding loop to reach the distant *TOX3* promoter, indicating that the role of *trans*-acting factors and the disruption of their binding site had an effect at very long range (Cowper-Sal lari *et al*, 2012).

Additional information also is required to characterise further the ZFN-modified clones, which may explain the divergent phenotypic appearances observed and identify additional mutations that might have occurred and that are not related to potential ZFN off-target effects. To this end, spectral karyotyping and SNP array (to assess copy number variation) should be conducted in each clone.

Isolating a true regulatory variant is complicated by linkage disequilibrium (LD) phenomena in the human genome. This is exemplified by studies of lactose intolerance, a common monogenic trait caused by *cis*-acting regulatory variants. Genetic studies in the Finnish population have shown a perfect correlation between the persistence of Lactase expression and the T allele of C/T-13910 variant (Enattah *et al*, 2002), further supported by *in vitro* studies showing functional differences between the alleles (Olds and Sibley, 2003). However, subsequent studies identified individuals who were heterozygous for the persistent allele but showed equal expression of *Lactase* alleles (Poulter *et al*, 2003), suggesting that this variant was actually in LD with the true causative regulatory variant.

6.2.2. FGFR2 expression and ASE

Additional results from the patient samples study did not establish clearly if the risk associated with rs2981578 is mediated via epithelial or mesenchymal cell types, with respect to which cell type was responsible for initiating or driving of tumour development. It would therefore be very interesting to look at the impact of rs2981578 in the tumour microenvironment, in particular in fibroblasts and myoepithelial cells (that support the epithelial cells in the breast lobules and ducts). The ZFN-genome editing process could be applied to mammary gland fibroblasts in order to create similar cell lines to those obtained for epithelial cells. The use of an organotypic model to study invasion could be extended to a 3D model of co-cultures of different cell types (epithelial and mesenchymal), as developed by Prof Louise Jones group in the Centre for Tumour Biology, at Barts Cancer Institute (Holliday *et al*, 2009). Ultimately, these cell lines could also be injected into immune-deficient mice to test the tumourigenicity of the cell models in an *in vivo* environment in the presence or absence of oestrogens. In parallel to the initial *in vitro* and animal studies, the levels of FGFR2 expression could be quantified by immunohistochemistry in an extended cohort of breast cancer patients.

6.2.3. GFP-tagged FGFR2 construct

The *FGFR2* ZFNs are very specific genomic scissors and their function cannot be extended to target any other areas of the genome. They can, however, be employed to introduce a transgene at the endogenous *FGFR2* locus. The ZFN cutting site, in the second intron, after the ATG start codon that initiates FGFR2 translation, would conveniently introduce that transgene under the control of the endogenous *FGFR2* promoter. This would constitute an obvious advantage compared to other strategies, since the *trans* and *cis* acting factors controlling *FGFR2* expression, for the most part, would still be present. A construct was created for such a purpose, containing the FGFR2-b isoform cDNA tagged with green fluorescent protein, the splicing acceptor sequence upstream of exon 3 of *FGFR2* (sA) and a neomycin resistance cassette (Appendix 10 and 11).

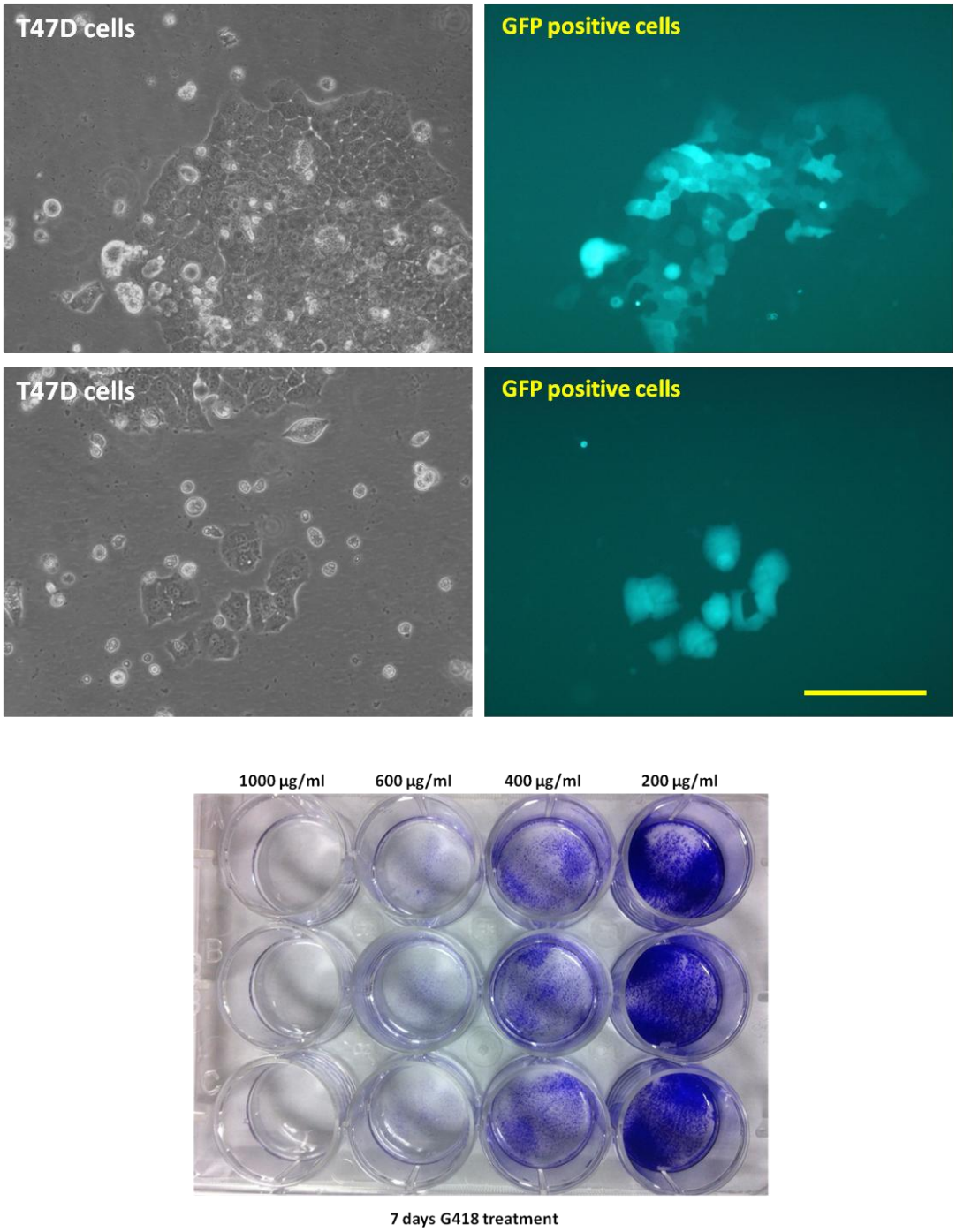


Figure 6.1: Endogenous expression of the FGFR2 tagged construct in T47D cells

T47D cells observed under transmission and UV light microscopy 19 days post transfection with ZFN mRNA and FGFR2-GFP-Neo construct (scale bar = 50 microns. The cells were cultured in medium containing 600 µg/ml of Geneticin (G418) from day 11 post-transfection (as estimated by a Kill curve experiment, bottom plate). G418 concentration required for Neomycin resistant cell selection. T47D cells were exposed to different concentrations of G418 (200 µg/ml to 1 mg/ml) over a period of one week. The surviving cells were stained with Coomassie blue to visualise the effect of the drug on this particular cell line. 600 µg/ml was the chosen concentration for antibiotic selection in the T47D cells.

A polyadenylation signal was also introduced after the GFP sequence in order to short-cut the expression of the endogenous *FGFR2*, and substitute the tagged protein instead (Appendix 10 and 11) (Gutschner *et al*, 2011).

Preliminary results showed that the FGFR2b-GFP/neo construct was successfully introduced and expressed in T47D cells treated with G418 drug (Fig. 6.1). T47D cells were chosen for this application, being one of the breast cancer cell lines expressing high levels of FGFR2 (as determined by publicly available expression microarray data (CCLE, 2012)). GFP expression was observed two weeks after ZFN mediated genome editing, however, the cell culture presented bacterial contamination before clones could be isolated. Although not included in the Results chapters, this was nevertheless a proof of concept that the tagged FGFR2 construct was successfully expressed by these cells. It remains to be demonstrated that the construct successfully integrated at the *FGFR2* endogenous locus. The applications for this construct are long ranging: from study of the cellular localisation of the receptor to the co-culture of cancer cells stably expressing either a green IIIb receptor isoform with other cells expressing a red (RFP) IIIc isoform. Another possibility is the modification of the wild-type receptor using Site Directed Mutagenesis to study the importance of known oncogenic or inherited mutations of the protein signalling cascade or cellular localisation under different conditions. For instance, S252W is an activating mutation, located in the acid box of the receptor, which is often present in FGFR2-dependent skeletal disorders and also is seen in breast and uterine cancer (Pollock *et al*, 2007). Because the ZFN editing method usually targets one allele of a gene at a time, we could also insert two different tagged constructs of FGFR2 isoforms (b and c) in the same individual diploid cell, at each endogenous locus, and observe their behaviour *in situ* using fluorescence microscopy. It may of course be possible that the tagged GFP alters the function and cellular localisation of FGFR2, by modifying its size or preventing adaptor proteins to bind to its intracellular tail. The bi-cistronic expression of FGFR2 and GFP could therefore be obtained by including an internal ribosome entry site (IRES) between the two cDNA sequences (Gurtu *et al*, 1996).

6.3. Conclusion

To conclude, ZFN-mediated genome editing showed the promising perspective of studying the role of risk SNP alleles in cancer cell lines. The collective data from this study showed that rs2981578 was not alone in mediating the breast cancer risk and that a combination of other SNP alleles might be required to confer a pre-carcinogenic state to the mammary gland cells via a mechanism involving FGFR2, ERα and the pioneer factor FOXA1.

CHAPTER 7

APPENDICES

7. Appendices

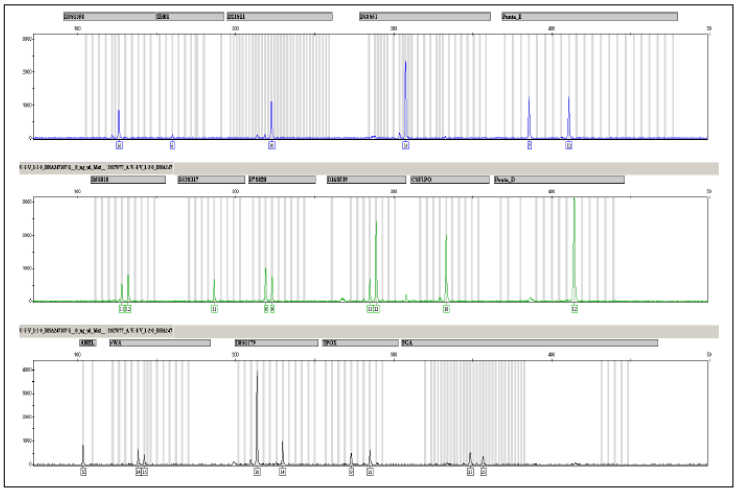
Appendix 1: STR profiling of MCF7 cells and spectral karyotyping

Cell Line Authentication Service – 16 Loci Service Results

Full descriptions of loci names can be found on page 4 of this document. This report is subject to LGC Standards terms and conditions. If you have any questions please contact the LGC Standards office +44 (0)208943 8489 or cell@lgcstandards.com.

1

Powerplex16 Loci	ATCC reference HTB-22	Customer sample MCF7
AMELO	X,X	X,X
D3		16,16
THO1	6,6	(6,6)
D21		30,30
D18		14,14
PentaE		7,12
D5	11,12	11,12
D13	11,11	11,11
D7	8,9	8,9
D16	11,12	11*,12
CSF	10,10	10,10
PentaD		12,12
VWA	14,15	14,15
D8		10,14*
TPOX	9,12	9,12
FGA		23,25

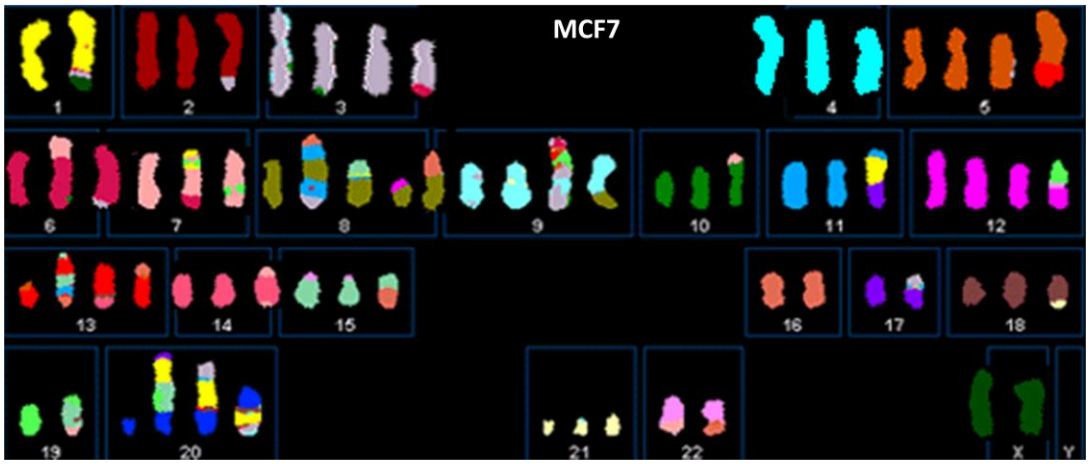


*Peak Area Difference
(n) Below-threshold peak
+ stutter

NOTE
ATCC use a 9 loci profile, here LGC Standards has profiled those same 9 plus 7 others.
Refer to page 5 for terms used.

MCF7
This profile matches all of the available 9 loci STR profile from ATCC for HTB-22.

Please note
Locus THO1 is below the recommended threshold for homozygous peaks.



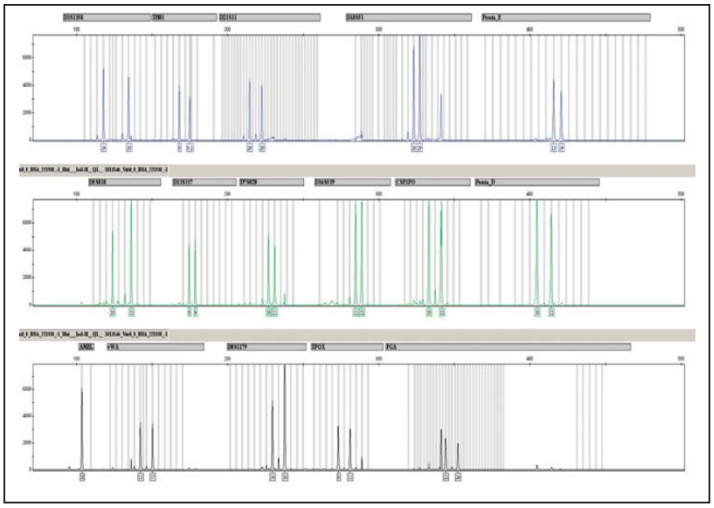
Appendix 2: STR profiling of MCF10A cells and spectral karyotyping

Cell Line Authentication Service – 16 Loci Service Results

Full descriptions of loci names can be found on page 4 of this document. This report is subject to LGC Standards terms and conditions. If you have any questions please contact the LGC Standards office +44 (0)208943 8489 or cell@lgcstandards.com.

12

Powerplex16 Loci	ATCC reference CRL-10317	Customer sample MCF10A
AMELO	X,X	X,X
D3		14,18
THO1	8,9.3	8,9.3
D21		28,30
D18		18,19
PentaE		13,14
D5	10,13	10,13
D13	8,9	8,9
D7	10,11	10,11
D16	11,12	11,12
CSF	10,12	10,12
PentaD		10,12
VWA	15,17	15,17
D8		14,16
TPOX	9,11	9,11
FGA		22,24



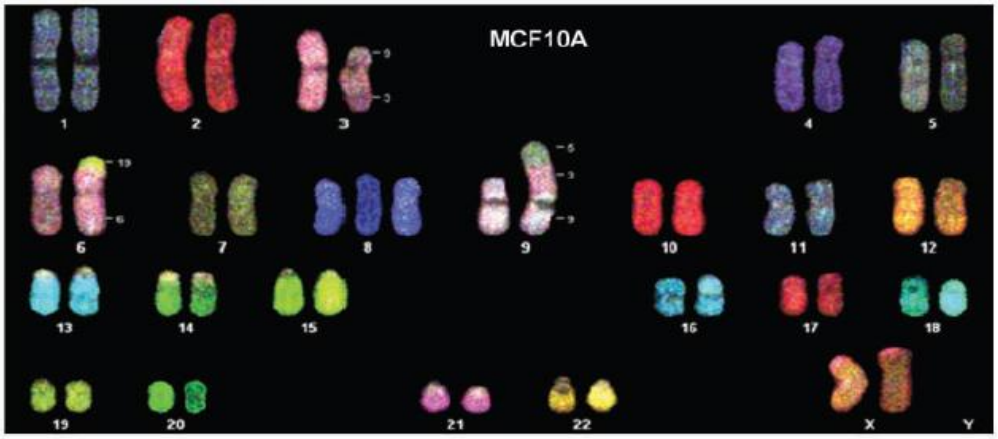
*Peak Area Difference
(n) Below-threshold peak
* stutter

NOTE

ATCC use a 9 loci profile, here LGC Standards has profiled those same 9 plus 7 others.
Refer to page 5 for terms used.

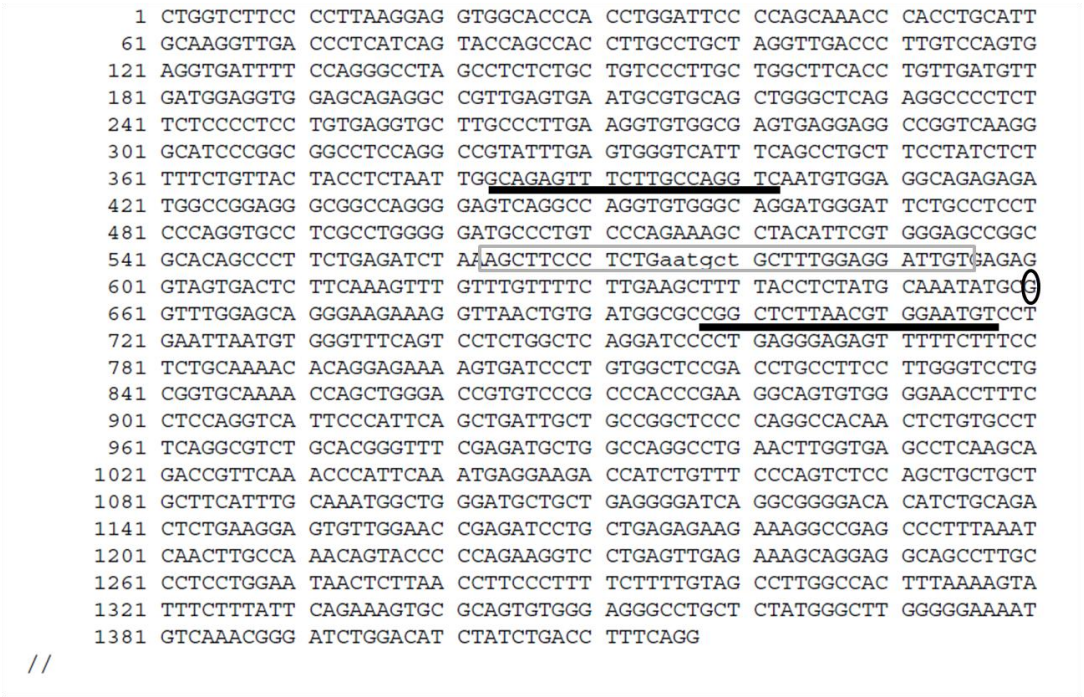
MCF10A

This profile matches all of the available 9 loci STR profile from ATCC for CRL-10317.



Appendix 3: ZFN primers and target site location relative to rs2981578

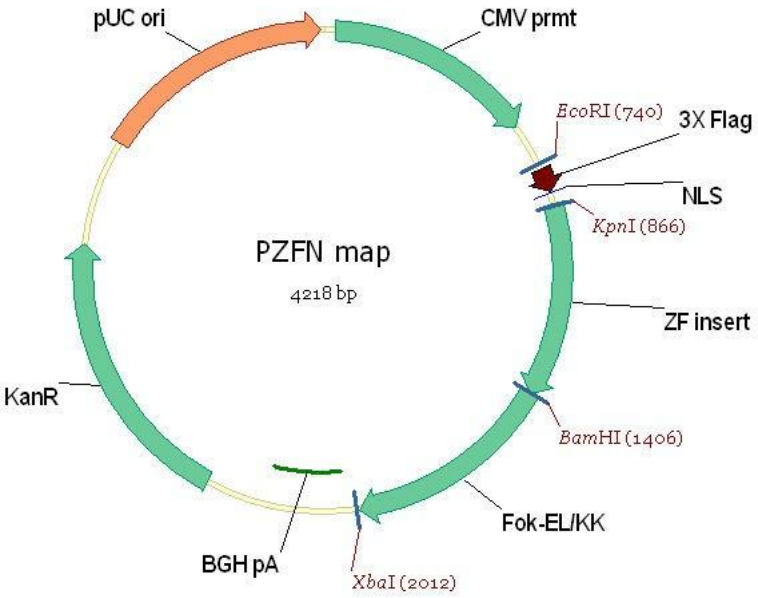
ZFN primers binding sites are underlined with a thick black line. The ZFN binding (upper case) and cutting (lower case) site is in the grey box. The SNP rs2981578 is circled in black.



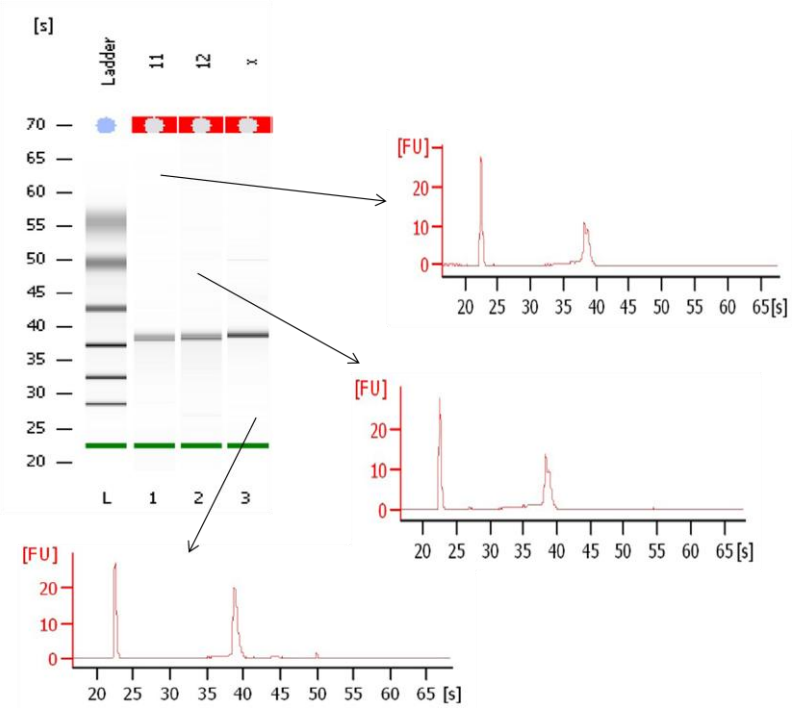
Appendix 4: Map of the ZFN plasmid CompoZr (Sigma) and quality of synthesized ZFN mRNA

A) The constructs, PZFN1 and PZFN2, containing both CMV and T7 promoters for use directly in eukaryotic cells and in an *in vitro* transcription reaction, respectively. All ZFNs are triple FLAG-peptide tagged (AspTyrLysAspAspAspAspLys) at the N-terminus, for protein detection by western blot. The restriction enzyme *Xba I* cuts just after the stop codon and was be used to linearise the template for mRNA production. The construct confers Kanamycin resistance to bacteria. B) NanoChIP RNA quality assay showing three different batch of synthesized ZFN mRNA.

A



B



Appendix 5: Centificate of Analysis CompoZr custom ZFN

SIGMA-ALDRICH®

sigma-aldrich.com

3050 Spruce Street, St. Louis, MO 63103 USA
Tel: (800) 521-8956 (314) 771-5765 Fax: (800) 325-5052 (314) 771-5757

Certificate of Analysis

Product Name: CompoZr™ Custom Zinc Finger Nucleases

Storage Temperature: ¯80°C

Product Number: CSTZFN-1KT

Product Brand: Sigma-Aldrich

CAS Number: None

Target Gene: FGFR2

Lot Number: 09250929MN

Kit Component	Test	Specification	Result
PZFN1	A260/A280 Ratio	1.8-2.0	1.9
PZFN1	Concentration	450-550 ng/ul	550ng/ul
PZFN2	A260/A280 Ratio	1.8-2.0	1.9
PZFN2	Concentration	450-550 ng/ul	511ng/ul
ZFN mRNA	A260/A280 Ratio	>1.8	2.2
PZFN1 & PZFN2	ZFN Activity in K562 Cells	>1%	1.9%
ZFN mRNA	ZFN Activity in K562 Cells	>1%	7.6%

Kit Component	Sequence – 5'-3'
ZFN Primer F	GCAGAGTTTCTTGCCAGGTC
ZFN Primer R	ACATTCCACGTTAAGAGCCG

Cel-1 PCR Parameter	Annealing Temperature = 57°C

3050 Spruce Street, St. Louis, MO 63103 USA
Tel: (800) 521-8956 (314) 771-5765 Fax: (800) 325-5052 (314) 771-5757

Certificate of Analysis

Product Name: CompoZr™ Custom Zinc Finger Nucleases

Storage Temperature: ¯80°C

Product Number: CSTZFN-1KT

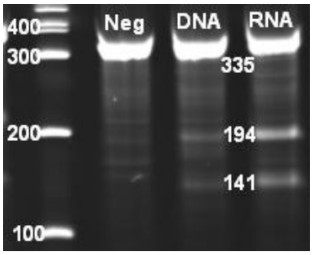
Product Brand: Sigma-Aldrich

CAS Number: None

Target Gene: FGFR2

Lot Number: 09250929MN

Surveyor Mutation Detection Assay Results Image



Zinc Finger Nuclease Binding/cutting site

AGCTTCCCTCTGaatgctGCTTTGGAGGATTGT

QC ACCEPTANCE DATE: November 12, 2009

Jessica Dillender

Functional Genomics

St. Louis, Missouri USA

SIGMA-ALDRICH®
sigma-aldrich.com
3050 Spruce Street, St. Louis, MO 63103 USA
Tel: (800) 521-8956 (314) 771-5765 Fax: (800) 325-5052 (314) 771-5757

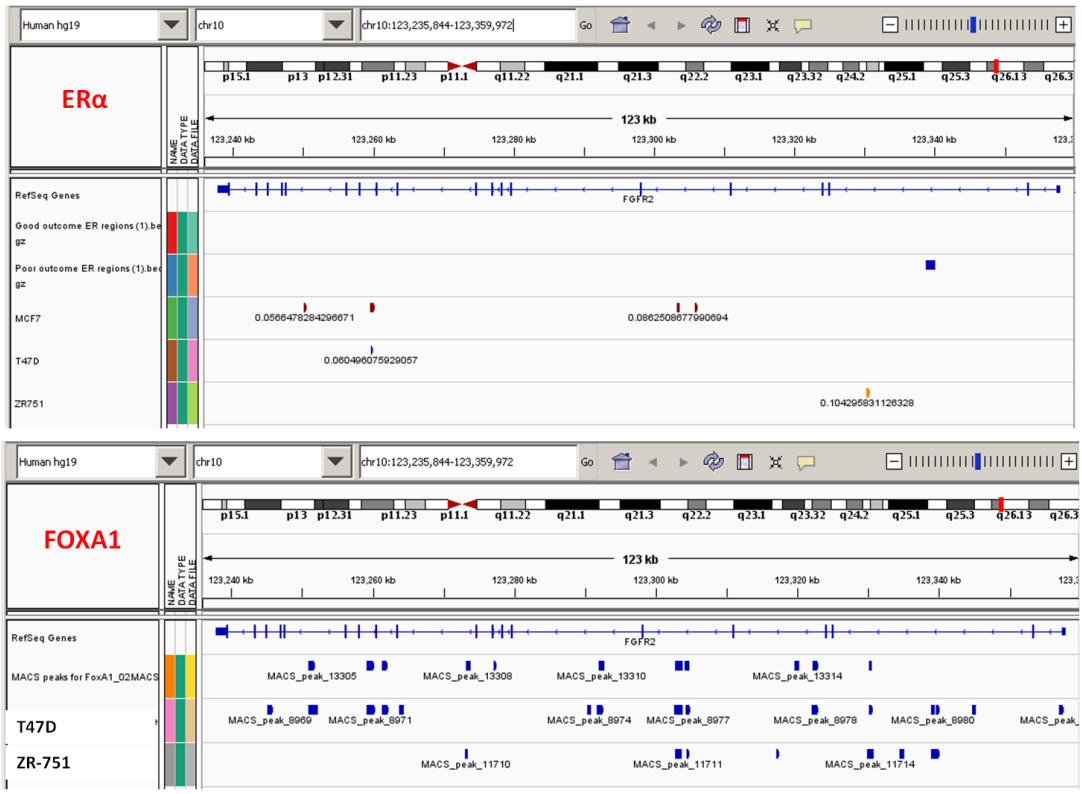
Certificate of Analysis

Product Description:

Kit Component Number	Format	Volume	Product Specifics
PZFN1	Plasmid DNA	45ul	ZFN Targeting Reverse Strand
PZFN2	Plasmid DNA	45ul	ZFN Targeting Forward Strand
ZFN mRNA	mRNA	5ul	Pooled mRNA for both ZFNs 10 Ready to use aliquots, 2ug of each ZFN per vial
ZFN Primers Forward and Reverse	Oligos	1 ml	25uM = working concentration for ZFN activity assay
CDZFN	Genomic Mammalian DNA	50ul	Control DNA for ZFN activity assay.

Appendix 6: ChIP-seq data for ERα and FOXA1 in breast cancer patients and breast cancer cell lines, at the *FGFR2* locus

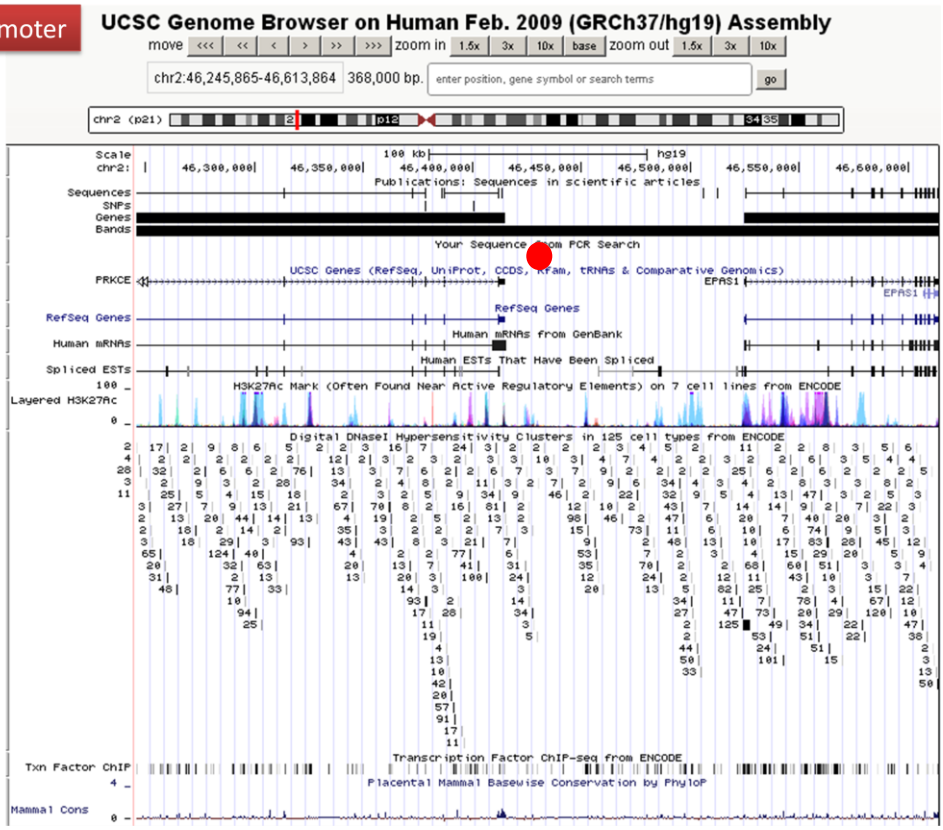
Data from (Ross-Innes *et al*, 2012) visualised using IGV2.2 software (Broad-Novartis Institute). FOXA1 and ERα ChIP seq data were available on the Carroll lab website (http://www.carroll-lab.org.uk/data). SNP rs2981578 is located on chr10:123330300 in the NCBI36/hg18 build of the human genome assembly.



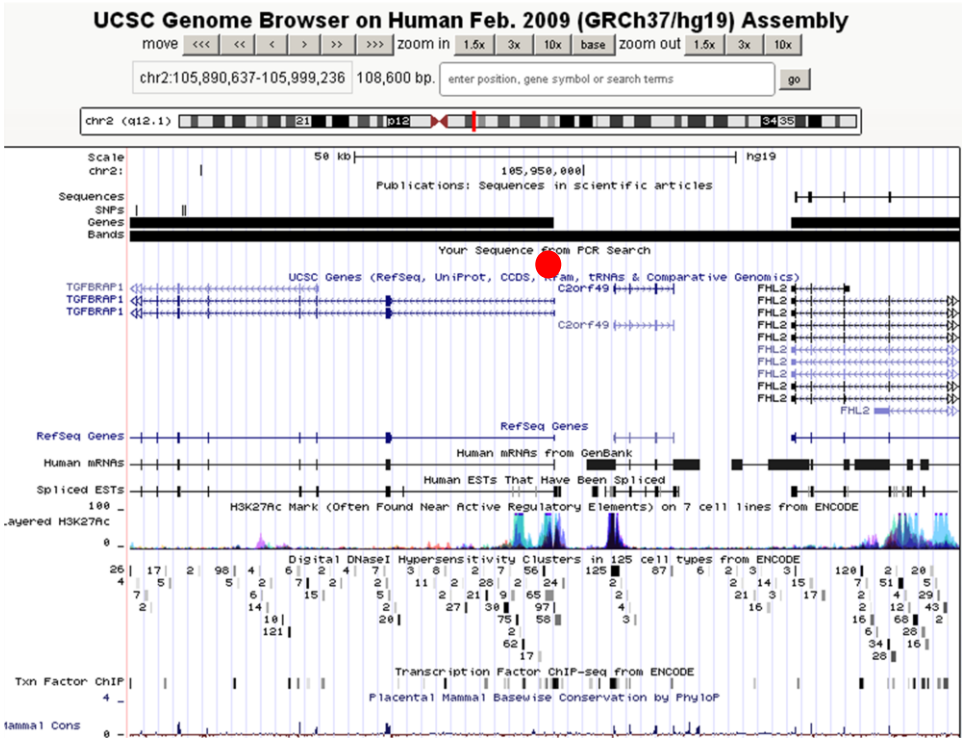
Appendix 7: FGFR2 ZFN off-target binding sites and their genomic context

Red dot represents the ZFN off-target binding site

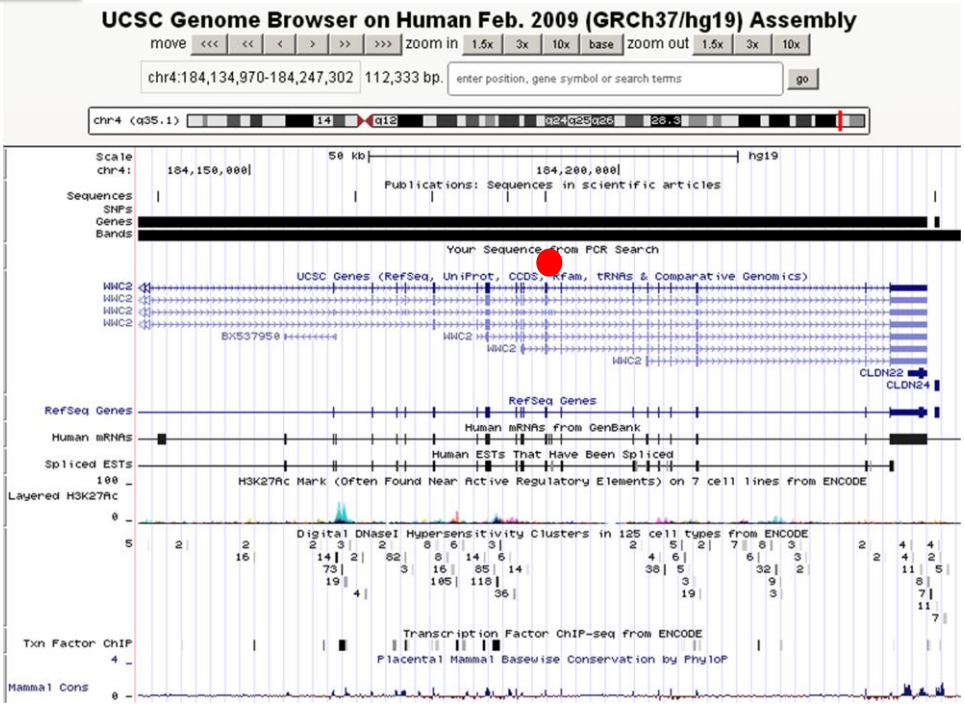
PRKCε promoter



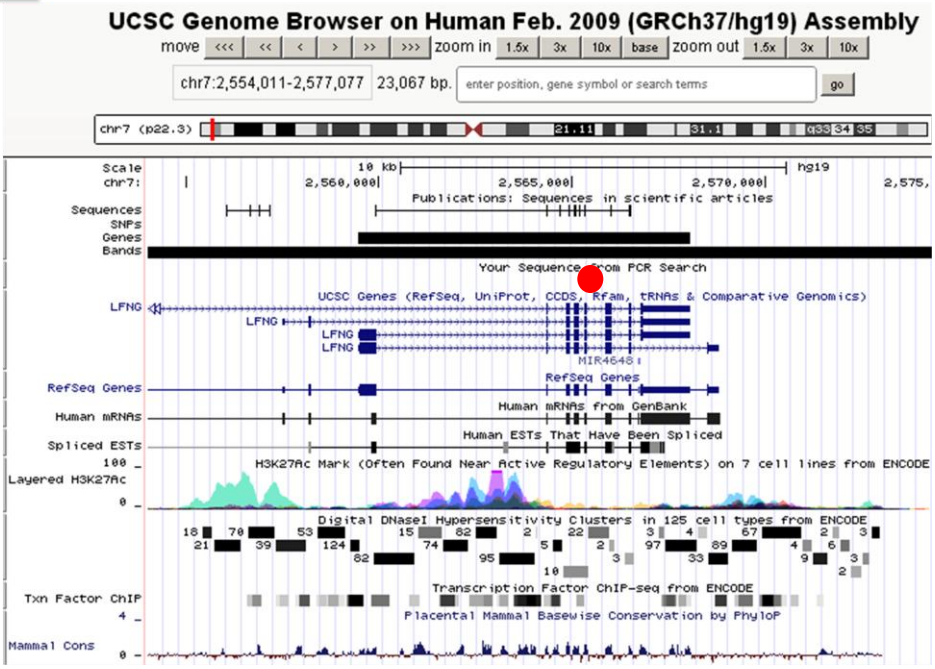
TGFβRAP1 promoter



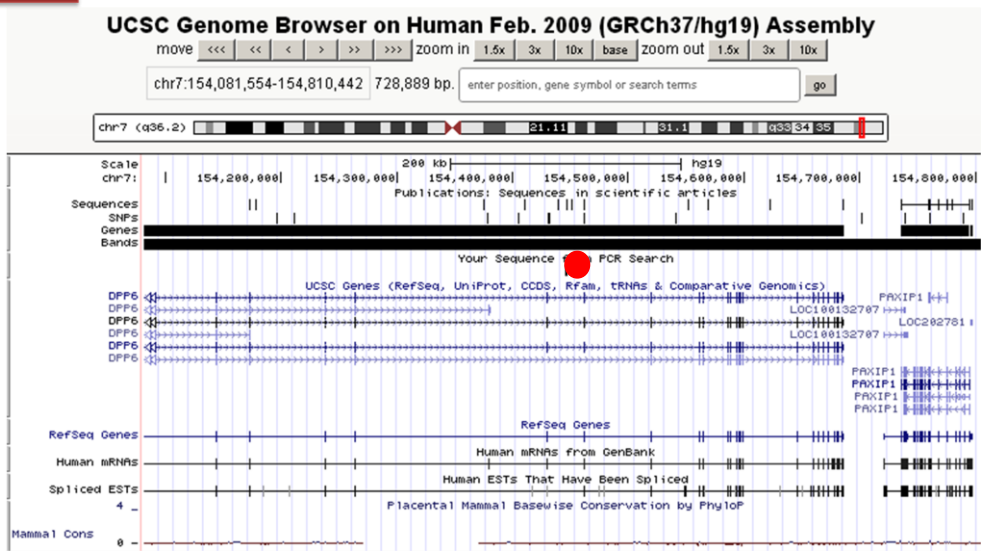
WWC2 intron



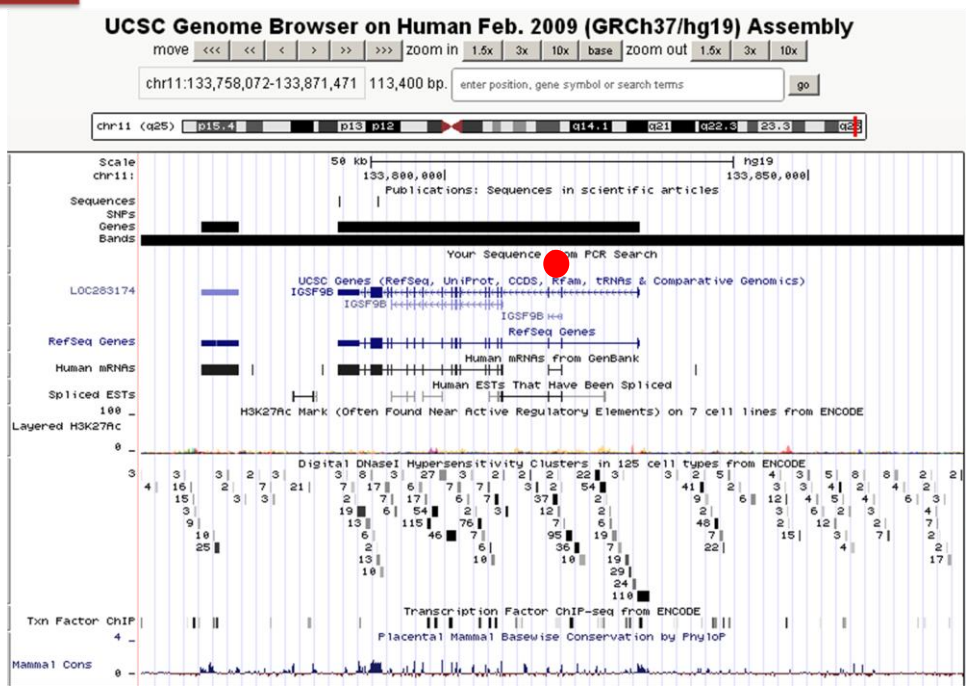
LFNG intron



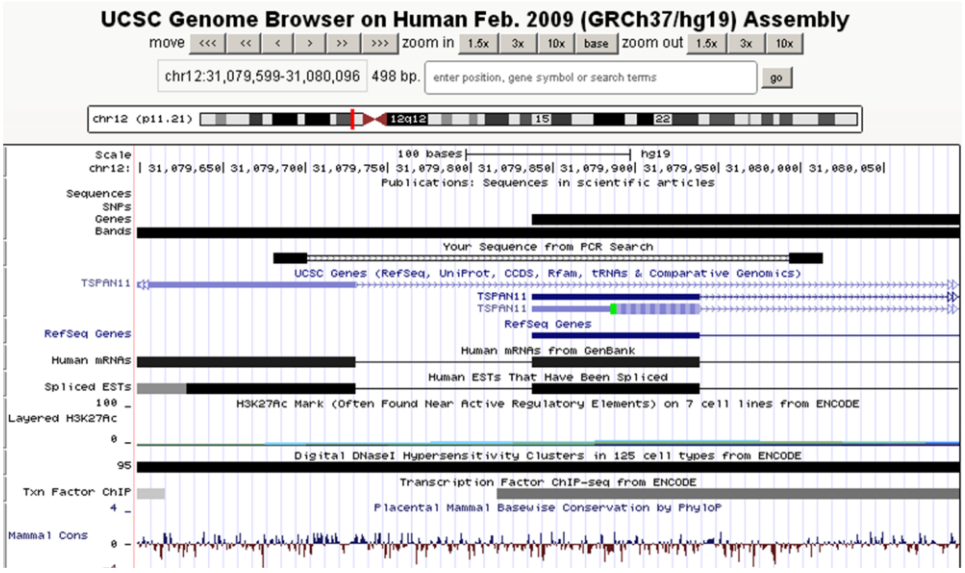
DPP6 intron



IGSF9B intron



TSPAN11 5'UTR



Appendix 8: Sequencing results of FGFR2 ZFN off-target loci

http://genome.ucsc.edu/cgi-bin/hgBlat

Matching bases in cDNA and genomic sequences are colored **blue** and capitalized. **Light blue** bases mark the boundaries of gaps in either sequence (often splice sites).

Intergenic 1 (*PRKCε* promoter)

E4 (Ctr 1)

```
ccttgggcca agctcaagca agtagcattc ctggaatctg ttgtttatta  46429669
gcagggatga gaggacagga caccacttgc tggttaaacc atctctagac  46429719
CTTATCTCTG cCCAGCACCT TTAAaCACAA ACTTTGCCCT GGGGCTGCAG  46429769
CCCTCTGAAG CCACCGGAGC TAAATGAGGT GCATTTGGGT TTTGAGCCTG  46429819
TTGATTCACA CTTATCTCCA GAGAGTTCTC CTTCCTCCCC TAAAGCTCCC  46429869
TCTAGAAACT TCCCTCAGGA CCCAAGAGGG AACCTCTGCC AAACCAGCTC  46429919
AACCTCCACC ATCTAATTCC ACTGGTATGG ATAAAGCTCA CATCATCCCA  46429969
GGCCTCAAAA CTCACAAGAC AAGGAGTTCA GATGGTAGTT TAAAATGTTT  46430019
CCCTTCTCCT AGCCCTGCCC AAGAGACTAA Ttccctgtta cttcagcctt  46430069
caccagaagc ccagagtcct ccagggtctt actgctctga tccatccagg  46430119
cctagtttta tctaaccccc tacctattcc c
```

F4 (Ctr 2)

```
gattgttcct tgggccaagc tcaagcaagt agcattcctg gaatctgttg  46429662
tttattagca gggatgagag gacaggacac cacttgctgg ttaaaccatc  46429712
TCTAgACCTT ATCTCTGCCC AGCACCTTta AACACAAACT TTGCCctGGG  46429762
GCTGCAgCCC TCTGAAGCCA CCGGAGCTAA aTGAgGTGCA TTTGGGTTTT  46429812
GAGCCTGTTG ATTCACACTT ATCTCCaGAG AGTTCTCCTT CCTCCCCTAA  46429862
AGCTCCCTCT AGAAACTTCC CTCAGgACCC AagagGGAAC CTCTGCCAAA  46429912
CCAGCTCAAC CTCCACCATC TAATTCCACT GGTATGGATA AAGCTCACAT  46429962
CATCCCAGGC CTCAAAACTC ACAAgaCAAG GAGTTCAGAT GGtagTTTAA  46430012
AATGTTTCCC TTCTCCTAGC CCTGCCCAaG AGACTAATtc cctgttactt  46430062
cagccttcac cagaagccca gagtcctcca gggtcttact gctctgatcc  46430112
atccaggcct agttttatct aacccccctac ctattccc
```

E11 (Ctr 3)

```
ccttgggcca agctcaagca agtagcattc ctggaatctg ttgtttatta  46429669
gcagggatga gaggacagga caccacttgc tggttaaacc atctctagac  46429719
CTTATCTCTG cCCAGCACCT TTAAaCACAA ACTTTGCCCT GGGGCTGCAg  46429769
CCCTCTGAAG CCACCGGAGC TAAATGAGGT GCATTTGGGT TTTGAGCCTG  46429819
TTGATTCACA CTTATCTCCa GAGAGTTCTC CTTCCTCCCC TAAAGCTCCC  46429869
TCTAGAAACT TCCCTCAGgA CCCAAgagGG AACCTCTGCC AAACCAGCTC  46429919
AACCTCCACC ATCTAATTCC ACTGGTATGG ATAAAGCTCA CATCATCCCA  46429969
GGCCTCAAAA CTCACAaGAC AAGGAGTTCA GATGGtAGTT TAAAATGTTT  46430019
CCCTTCTCCT AGCCCTGCCC AAGAGACTAA Ttccctgtta cttcagcctt  46430069
caccagaagc ccagagtcct ccagggtctt actgctctga tccatccagg  46430119
cctagtttta tctaaccccc tacctattcc c
```

A8 (Het 1)

```
ccttgggcca agctcaagca agtagcattc ctggaatctg ttgtttatta  46429669
gcagggatga gaggacagga caccacttgc tggttaaacc atctctagac  46429719
CTTATCTCTG cCCAGCACCT TTAAACACAA ACTTTGCCCT GGGGCTGCAg  46429769
CCCTCTGAAG CCACCGGAGC TAAATGAGGT GCATTTGGGT TTTGAGCCTG  46429819
TTGATTCACA CTTATCTCCa GAGAGTTCTC CTTCCTCCCC TAAAGCTCCC  46429869
TCTAGAAACT TCCCTCAGGA CCCAAgAGGG AACCTCTGCC AAACCAGCTC  46429919
AACCTCCACC ATCTAATTCC ACTGGTATGG ATAAAGCTCA CATCATCCCA  46429969
GGCCTCAAAA CTCACAAGAC AAGGAGTTCA GATGGTAGTT TAAAATGTTT  46430019
CCCTTCTCCT AGCCCTGCCC AAGAGACTAA Ttccctgtta cttcagcctt  46430069
```

caccagaagc ccagagtcct ccagggtctt actgctctga tccatccagg 46430119
cctagtttta tctaaccccc tacctattcc c

C4 (Het 2)
ccttgggcca agctcaagca agtagcattc ctggaatctg ttgtttatta 46429669
gcagggatga gaggacagga caccacttgc tggttaaacc atctctagac 46429719
CTTATCTCTG cCCAGCACCT TTAAaCACAA ACTTTGCCCT GGGGCTGCAG 46429769
CCCTCTGAAG CCACCGGAGC TAAATGAGGT GCATTTGGGT TTTGAGCCTG 46429819
TTGATTCACA CTTATCTCCa GAGAGTTCTC CTTCCTCCCC TAAAGCTCCC 46429869
TCTAGAAACT TCCCTCAGGA CCCAAgagGG AACCTCTGCC AAACCAGCTC 46429919
AACCTCCACC ATCTAATTCC ACTGGTATGG ATAAAGCTCA CATCATCCCA 46429969
GGCCTCAAAA CTCACAAGAC AAGGAGTTCA GATGGTAgTT TAAAATGTTT 46430019
CCCTTCTCCT AGCCCTGCCC AAGAGACTAA Ttccctgtta cttcagcctt 46430069
caccagaagc ccagagtcct ccagggtctt actgctctga tccatccagg 46430119
cctagtttta tctaaccccc tacctattcc c

G11 (Het 3)
ccttgggcca agctcaagca agtagcattc ctggaatctg ttgtttatta 46429669
gcagggatga gaggacagga caccacttgc tggttaaacc atctctagac 46429719
CTTATCTCTG CCcAGCACCT TTaAACACAA ACTTTGCCCT GGGGCTGCAG 46429769
CCCTCTGAAG CCACCGGAGC TAAATGAGGT GCATTTGGGT TTTGAGCCTG 46429819
TTGATTCACA CTTATCTCCa GAGAGTTCTC CTTCCTCCCC TAAAGCTCCC 46429869
TCTAGAAACT TCCCTCAGGA CCCAAGAGGG AACCTCTGCC AAACCAGCTC 46429919
AACCTCCACC ATCTAATTCC ACTGGTATGG ATAAAGCTCA CATCATCCCA 46429969
GGCCTCAAAA CTCACAAGAC AAGGAGTTCA GATGGTAGTT TAAAATGTTT 46430019
CCCTTCTCCT AGCCCTGCCC AAGAGACTAA Ttccctgtta cttcagcctt 46430069
caccagaagc ccagagtcct ccagggtctt actgctctga tccatccagg 46430119
cctagtttta tctaaccccc tacctattcc c

Intergenic 2 (*TGFβRAP1* promoter)

E11(Ctr3)
gtttaatcac tgatgaagtg ctcatgggca agcaccttgt ctatcaggaa 105945134
ggaaggcagc tgcgggtgtg attgtggtgg caagttaagg cactgagaag 105945084
GTGTGCAGGG aAGATCCTAC TGCGTTGTGT GTGTTGGCTT GGTTGGAGTC 105945034
AGGATGAACG AATGTAGAGG GACTGTGTGG TGGCTGCTAC TCTCTCATGG 105944984
TGGACATGCT TCTCTCTGTT GGTCAGCCGG AAGCTGGCAg ACAACTGAgA 105944934
GCTCGCAGTT CCaTGGTcag tcagAAGGAA CATTTGgtCA gCTGCTTCCT 105944884
TtTCTTTTTC CTTCCATTTC TGCCagcCgA gcaGCCTCAA CaGtCaTTGC 105944834
AgCTGCTTCT GTgATGTGTA GCAAAacgct gcctggagta ctagaaagtc 105944784
tactggctgg AATCTGCTGG AACCTGTCca taactagcct gccccaggtc 105944734
atctatgaaa tgggagtgat gctctgtggt ggtaatgaat gtggagtgtg 105944684
ggttcaaatc ctgcctctgc catttcct

F4 (Ctr2)
gtttaatcac tgatgaagtg ctcatgggca agcaccttgt ctatcaggaa 105945134
ggaaggcagc tgcgggtgtg attgtggtgg caagttaagg cactgagaag 105945084
GTGTGCAGGg aAGATCCTAC TGCGTTGTGT GTGTTGGCTT gGTTGGAGTC 105945034
AGGATGAACG AATGTAGAGG GACTGTGTGG TGGCTGCTAC TCTCTCATGG 105944984
TgGACATGCT TCTCTCTGTT GGTCAgCCGG AAGCTGGCAg ACAACTGAgA 105944934
GCTCGCAGTT CCaTGGTcag tcagAAGGAA CAtTTGgtCA gCTGCTTCCt 105944884
TTTCTTTTTC CTTCCATTTC TGCCagcCgA gcaGCCTCAA CAGtCaTTGc 105944834
agCTGCTTCT GTgATGTGTA gCAAAacgct gcctggagta ctagaaagtc 105944784
tactggctgg aatctgctgg aacctgtcca taactagcct gccccaggtc 105944734
atctatgaaa tgggagtgat gctct

E4 (Ctr 1)
ttaatcactg atgaagtgct catgggcaag caccttgtct atcaggaagg 105945132
aaggcagctg cgggtgtgat tgtggtggca agttaaggca ctgagaaggt 105945082
GTGCAGGGAa gatCCTACTG CGTTGTGTGT GTTGGCTTgG TTGGAGTCAG 105945032
GATGAACGAA TGTAGAGGGA CTGTGTGGTG GCTGCTACTC TCTCATGGTG 105944982
GACATGCTTC TCTCTGTTGG TCAgCCGGAA GCTGGCAgAC AACTGAgAGC 105944932
TCGCAgTTCC ATGGTCAgTC AgAAGGAACA TTTGGTCagC TGCTTCCtTT 105944882
TCTTTTTCCT TCCATTTCTG CCAgccgagc AGCCTCAACA GtCATTGCAg 105944832
CTGCTTCTGT GATGTGTAgC AAAACGCTGC CTGGAgTACT AgAAAgTCTA 105944782
CTGGctggaa tctgctggaa cctgtccata actagcctgc cccaggtcat 105944732
ctatgaaatg ggagtgatgc tctgtggtgg taatgaatgt ggagtgtggg 105944682
ttca

C4 (Het 2)
ttaatcactg atgaagtgct catgggcaag caccttgtct atcaggaagg 105945132
aaggcagctg cgggtgtgat tgtggtggca agttaaggca ctgagaaggt 105945082
GTGCAGGGaA GATCCTACTG CGTTGTGTGT GTTGGCTTGG TTGGAGTCAG 105945032
GATGAACGAA TGTAGAGGGA CTGTGTGGTG GCTGCTACTC TCTCATGGTG 105944982
GACATGCTTC TCTCTGTTGG TCAGCCGGAA GCTGGCAGAC AACTGAgAGC 105944932
TCGCAGTTCC ATGGtCAgTC AGAAGgAACA TTTGGTCAgC TGCTTCCTTT 105944882
TCTTTTTCCT TCCATTTCTG CCagCCgAgc AGCCTCAACA GtCaTTGCAG 105944832
CTGCTTCTGT GATGTGTAgC AAAACGCTGC CTGGAgTACT AgAAAgTCTA 105944782
CTGgCTGgAA TCTGCTGGAA CCTGTCcata actagcctgc cccaggtcat 105944732
ctatgaaatg ggagtgatgc tctgtggtgg taatgaatgt ggagtgtggg 105944682
ttcaaatcct gcctctgcca tttcct

A8 (Het 1)
agtgctcatg ggcaagcacc ttgtctatca ggaaggaagg cagctgcggg 105945118
tgtgattgtg gtggcaagtt aaggcactga gaaggtgtgc agggaagatc 105945068
CTACTGCGTT GTGTGTGTTG GCTTgGTTGG AGTCAGGATG AACGAATGTA 105945018
GAGGGACTGT GTGGTGGCTG CTACTCTCTC ATGGTGGACA TGCTTCTCTC 105944968
TGTTGGTCAg CCGGAAGCTG GCAgACAACT GAgaGCTCGC AGTTCCaTGG 105944918
Tcagtcag AA GGAACATTTG gtCAgCTGCT TCCTTTTCTT TTTCCTTCCA 105944868
TTTCTGCCag cCgAgcaGCC TCAACaGtCa TtgcagCTGC TTCTGTgATG 105944818
TGTAgCAAAa cgctgcctgg agtactagaa agTCTACTGg cTggAATCTG 105944768
CTGGAACCTG TCcataacta gcctgcccca ggtcatctat gaaatgggag 105944718
tgatgctctg tggtggtaat gaatgtggag tgtgggttca aatcctgcct 105944668
ctgccatttc ct

G11 (Het 3)
ttaatcactg atgaagtgct catgggcaag caccttgtct atcaggaagg 105945132
aaggcagctg cgggtgtgat tgtggtggca agttaaggca ctgagaaggt 105945082
GTGCAGGgaA GATCCTACTG CGTTGTGTGT GTTGGCTTGg tTgGAGTCAG 105945032
GATGAACGAA TGTAGAGGGA CTGTGTGGTG GCTGCTACTC TCTCATGGTg 105944982
GACATGCTTC TCTCTGTTGG TCAgCCGGAA GCTGGCAgAC AACTGAgAGC 105944932
TCGCAGTTCC ATGGtcagtc agAAGGAACA TTTGgtcagC TGCTTCCTTT 105944882
TCTTTTTCCT TCCATTTCTG CCAgcCgAgc AGCCTCAACA GtCATTGCAg 105944832
CTGCTTCTGT GATGTGTAgC AAAACgctgc ctggagtact agaaagTCTA 105944782
CTGgcTGgAA TCTGCTGGAA CCTGTCcata actagcctgc cccaggtcat 105944732
ctatgaaatg ggagtgatgc tctgtggtgg taatgaatgt ggagtgtggg 105944682
ttcaaatcct gcctctgcca tttcct

WWC2

E4 (Ctr 1)
attctcctcc accacagtat tggagctttc gggaagatgt gatgttactg 184190947
tttaaagcaa tatgacattt aaatgctaca gcagaagact tcacagttaa 184190997
CTAAtTGTGA GTttaATACa CTGTTGTGCG TaATAACCAg AAAACCATTA 184191047
TGTCCCAGTA AAGTTACATA AgTATTcaaa tgcAGGCATC TAgAGATGTG 184191097
CCATGTGTTC AgAAAACAAT GTGAGTCTCC AATTGAgCTT TTCTCTGcag 184191147
TCCAgTGGGA AGCACAgAAA CACATGTTTC ATGATGAACA AAGTTTAAgA 184191197
GGGGtAGGTT TCTTCTTAAT TTTTCTTCTT GTTTTCTTTA CTTGAAAAAA 184191247

ATATCTGATG ATaTTGTCTG ACAATAAAGT TAGAAAGAAG CAgagctaac 184191297
agactcgtta ggctatcaga aaggctttag tatcatagat gtcatattta 184191347
gtataataga tgtcagaggc tcccacaaag ttcataaata tt

F4 (Ctr 2)
cctccaccac agtattggag ctttcgggaa gatgtgatgt tactgtttaa 184190952
agcaatatga catttaaatg ctacagcaga agacttcaca gttaactaat 184191002
TGTGAGTTTa ATACACTGTT GTGCGTAATA ACCAgAAAAC CATTATGTCC 184191052
CAGTAAAGTT ACATAAGTAT TCAAATGCAG GCATCTAgAG ATGTGCCATG 184191102
TGTTCAgAAA ACAATGTGAG TCTCCAATTG AGCTTTTCTC TGCAGTCCAG 184191152
TGGGAAGCAC AgAAACACAT GTTTCATGAT GAACAAAGTT TAAgAGGGGT 184191202
AGGTTTCTTC TTAATTTTTC TTCTTGTTTT CTTTACTTGA AAAAAATATC 184191252
TgATGATaTT GTCTGACAAT AAAGTTAgaa agaagcagag ctaacagact 184191302
cgttaggcta tcagaaaggc tttagtatca tagatgtcat atttagtata 184191352
atagatgtca gaggctccca caaagtt

E11 (Ctr 3)
cctccaccac agtattggag ctttcgggaa gatgtgatgt tactgtttaa 184190952
agcaatatga catttaaatg ctacagcaga agacttcaca gttaactaat 184191002
TGTGAGTTta ATACaCTGTT GTGCGTaATA ACCAgAAAAC CATTATGTCC 184191052
CAGTAAAGTT ACATAAgTAT TCAAATGCAG GCaTCTAgAG ATGTGCCATG 184191102
TGTTCAgAAA acAATGTGAG TCTCCAATTG AgCTTTTCTC TGcagTCCAG 184191152
TGGGAAGCAC AgAAACACAT GTTTCATGAT GAACAAAGTT TAAgAGGGGT 184191202
AGGTTTCTTC TTAATTTTTC TTCTTGTTTT CTTTACTTGA AAAAAATATC 184191252
TGATGATaTT GTCTGACAAT AAAGTTAgAA AGAAGCAgag CTAACAgACT 184191302
CGTtaggcta tcagaaaggc tttagtatca tagatgtcat atttagtata 184191352
atagatgtca gaggctccca caaagttcat aaatattata ttgatagcct 184191402
taa

A8 (Het 1)
cctccaccac agtattggag ctttcgggaa gatgtgatgt tactgtttaa 184190952
agcaatatga catttaaatg ctacagcaga agacttcaca gttaactaat 184191002
TGTGAGTTta ATACaCTGTT GTGCGTaaTA ACCAgAAAAC CATTATGTCC 184191052
CAGTAAAGTT ACATAAGTAT TcaAAtGCAG GCaTCTAGAG ATGTGCCATG 184191102
TGTTCAgAAA ACAATGTGaG TCTCCAATTG AGCTTTTCTC TGCAGTCCAG 184191152
TGGGAAGCAC AgAAACACAT GTTTCATGAT GAACAAAGTT TAAgAGGGGT 184191202
AGGTTTCTTC TTAATTTTTC TTCTTGTTTT CTTTACTTGA AAAAAATATC 184191252
TgATGATaTT GTCTGACAAT AAAGTTAgaa agaagcagag ctaacagact 184191302
cgttaggcta tcagaaaggc tttagtatca tagatgtcat atttagtata 184191352
atagatgtca gaggctccca caaagtt

C4 (Het 2)
cctccaccac agtattggag ctttcgggaa gatgtgatgt tactgtttaa 184190952
agcaatatga catttaaatg ctacagcaga agacttcaca gttaactaat 184191002
TGTGAGTTta ATACACTGTT GTGCGTAATA ACCAgAAAAC CATTATGTCC 184191052
CAGTAAAGTT ACATAAGTAT TCAAATGCAG GCaTCTAgAG ATGTGCCATG 184191102
TGTTCAgAAA ACAATGTGAG TCTCCAATTG AGCTTTTCTC TGCAGTCCAG 184191152
TGGGAAGCAC AgAAACACAT GTTTCATGAT GAACAAAGTT TAAgAGGGGT 184191202
AGGTTTCTTC TTAATTTTTC TTCTTGTTTT CTTTACTTGA AAAAAATATC 184191252
TgATGATaTT GTCTGACAAT AAAGTTAgaa agaagcagag ctaacagact 184191302
cgttaggcta tcagaaaggc tttagtatca tagatgtcat atttagtata 184191352
atagatgtca gaggctccca caaagt

G11 (Het 3)
cctccaccac agtattggag ctttcgggaa gatgtgatgt tactgtttaa 184190952
agcaatatga catttaaatg ctacagcaga agacttcaca gttaactaat 184191002
TGTGAGTTta ATACACTGTT GTGCGTAATA ACCAgAAAAC CATTATGTCC 184191052
CAGTAAAGTT ACATAAGTAT TCAAATGCAG GCaTCTAgAG ATGTGCCATG 184191102
TGTTCAgAAA ACAATGTGAG TCTCCAATTG AGCTTTTCTC TGCAGTCCAG 184191152
tGGGAAGCAC AgAAACACAT GTTTCATGAT GAACAAAGTT TAAgAGGGGT 184191202
AGGTTTCTTC TTAATTTTTC TTCTTGTTTT CTTTACTTGA AAAAAATATC 184191252
TgATGATaTT GTCTGACAAT AAAGTTAgaa agaagcagag ctaacagact 184191302
cgttaggcta tcagaaaggc tttagtatca tagatgtcat atttagtata 184191352

atagatgtca gaggctccca caaagtt

***LFNG* (intron)**

E11 (Ctr 3)

```
cccttctccc agcgtcctgt ccacttctgg tttgccacgg gcggcgctgg  2565356
cttctgcatc agccgtgggc tggctctgaa gatgagcccg tgggccaggt  2565406
GAGTGCCctg caCAGGTTAG GCCAGCCCGG TCCCAGGCTC CTCGCCACTG  2565456
TGGGGCCTGG CTTAGTTCAT CTTCCCAGCC ATGGGGTGTC CCCAGCCTCC  2565506
TGTGTGGcAC TGCCCACTTA CTTCCTATAT TCCACTTCCC TCTgGGTTTC  2565556
AgAGGGCAGc TGTGTTTACG gCGGCTGCCC CCAAGcCTGA CCTGCTCAGA  2565606
GcAGcCAgGG GGGCGATGAg CACCCCAGGC ACCATCCggc aGGACTCTTC  2565656
CCTGcaCCCA gATTCCCTCC ACAgAGAGcC ACGGAGcACA gGAGCTGTGc  2565706
AGGGAGTGTG ccctggctgt ggccagggga ggcagaggga gctgcagccc  2565756
agagctctcc tcagggctcc tctccctgag gagtgcagcg cctttgcctg  2565806
gtggggcctc
```

F4 (Ctr 2)

```
cgtcctgtcc acttctggtt tgccacgggc ggcgctggct tctgcatcag  2565368
ccgtgggctg gctctgaaga tgagcccgtg ggccaggtga gtgccctgca  2565418
CAGGTTAGGC CAGCCCGGTC CCAGGCTCCT CGCCACTGTG GGGCCTGGCT  2565468
TAGTTCATcT TCCCAGCCAT GGGGTGtCCC CAGCCTCCTG TGTGGcACTG  2565518
CCCACTTACT TCCTATATTC CACTTCCCTC TGGGTTTCAg AGGGCAGcTG  2565568
TGTTTACgGC GGCTGCCCCC AAGcCTGACC TGCTCAGAGc AGcCAgGGGG  2565618
GCGATGAGCA CCCCAGGCAC CATCCGGCAG GACTCTTCCC TGCACCCAgA  2565668
TTCCCTCCAC AgAGAGCCAC GgAGCACAGg AGCTGTGCAG GgAGTGTGcc  2565718
ctggctgtgg ccaggggagg cagagggagc tgcagcccag agctctcctc  2565768
agggctcctc tccctgagga gtgcagcgcc tttgcctggt ggggcctc
```

E4 (Ctr 1)

```
cccagcgtcc tgtccacttc tggtttgcca cgggcggcgc tggcttctgc  2565363
atcagccgtg ggctggctct gaagatgagc ccgtgggcca ggtgagtgcc  2565413
CTGCAcAGGT TAGGCCAGCC CGGTCCCAGG CTCCTCGCCA CTGTGGGGCC  2565463
TGGCTTAgTT CATCTTCCCA GCCATGGGGT GTCCCCAGCC TCCTGTGTGG  2565513
CACTGCCCAC TTACTTCCTA TATTCCACTT CCCTCTGGGT TTCAgAGGGC  2565563
AGcTGTGTTT ACGgCGGCTG CCCCCAAGcC TGACCTGCTC AGAGcagcca  2565613
gGGGGGCGAT GAgCACCCCA GGCACCATCC GGcaGGACTC TTCCCTGCAC  2565663
CCAgATTCCC TCCACAgaga gccacggagc acaggagctg tgcagggagt  2565713
gtgccctggc tgtggccagg ggaggcagag ggagctgcag cccagagctc  2565763
tcctcagggc tcctct
```

C4 (Het 2)

```
ccttctccca gcgtcctgtc cacttctggt ttgccacggg cggcgctggc  2565357
ttctgcatca gccgtgggct ggctctgaag atgagcccgt gggccaggtg  2565407
AGTGcCCTGC AcAGGTTAGG CCAGCCCGGT CCcAGGCTCC TCGCCACTGT  2565457
GGGGCCTGGC TTAgTTCATC TTCCCAGCCA TGGGGTGTCC CCAGCCTCCT  2565507
GTGTGGCACT GCCCACTTAC TTCCTATATT CCACTTCCCT CTGGGTTTCA  2565557
gAGGGCAgCT GTGTTTACgG CGGCTGCCCC CAAGCCTGAC CTGCTCAGAG  2565607
cAGCCAGGGG GgCGATGAgC ACCCCAGGCA CCATCCgGCA GGACTCTTCC  2565657
CTGCACCCAg atTCCCTCCA CAgagagcca cgGAGCACAG gagCTGTGCa  2565707
GGGAGTGTGc cctggctgtg gccaggggag gcagagggag ctgcagccca  2565757
gagctctcct cagggctcct ctccctgagg agtgcagcgc ctttgcctgg  2565807
tggggcctc
```

A8 (Het 1)

```
cccttctccc agcgtcctgt ccacttctgg tttgccacgg gcggcgctgg  2565356
cttctgcatc agccgtgggc tggctctgaa gatgagcccg tgggccaggt  2565406
GAGTGcCCTG CAcAGGTTAG GCCAGCCCGG TCCCAGGCTC CTCGCCACTG  2565456
```

TGGGGCCTGG CTTAgTTCAT CTTCCCAGCC ATGGGGTGTC CCCAGCCTCC 2565506
TGTGTGGcAC TGCCCACTTA CTTCCTATAT TCCACTTCCC TCTGGGTTTC 2565556
AgAGGGCAGc TGTGTTTACG gCGGCTGCCC CCaAGcCTGA CCTGCTCAGA 2565606
GcAGcCAgGG GGGCGATGAg CACCCCAGGC ACCATCCGGc aGGACTCTTC 2565656
CCTGCACCCA gATTCCCTCC ACAgagagcc acggagcaca ggagctgtgc 2565706
agggagtgtg ccctggctgt ggccagggga ggcagaggga gctgcagccc 2565756
agagctctcc tcagggctcc tct

G11 (Het 3)
cccagcgtcc tgtccacttc tggtttgcca cgggcggcgc tggcttctgc 2565363
atcagccgtg ggctggctct gaagatgagc ccgtgggcca ggtgagtgcc 2565413
CTGCAcAGGT TAGGCCAGCC CGGTCCcAGG CTCCTCGCCA CTGTGGGGCC 2565463
TGGCTTAgTT CATCTTCCCA GCCATGGGGT GTCCCCAGCC TCCTGTGTGG 2565513
cACTGCCCAC TTACTTCCTA TATTCCACTT CCCTCTGGGT TTCAgAGGGC 2565563
AgCTGTGTTT ACgGCGGCTG CCCCCAAGCC TGACCTGCTC AGAGcAGCCa 2565613
gGGGGGCGAT GAgCACCCCA GGCACCATCC GGcAGGACTC TTCCCTGCAC 2565663
CCAgATTCCC TCCACAgaga gccacgGAGC ACAGgagctg tgcagGGAGT 2565713
GTGccctggc tgtggccagg ggaggcagag ggagctgcag cccagagctc 2565763
tcctcagggc tcctctccct gaggagtgca gcgcctttgc ctggtggggc 2565813
ctc

***DPP6* (intron)**

E11 (Ctr 3)
tttacattag caattacatt gaattaggga ttataagtag tctagaggtg 154445827
gcttaaagga tacgggagga tgtgcttaca ttatatgcaa ataccacaca 154445877
TTTTATATca AGGACTTGAG CATCTGTGGA TTTTGGTATC TGCAGGGGTG 154445927
TCCCGGAACC AATCTTCCAT GGATACCAAG GATGACTGTG CTCATATTTG 154445977
TGATCATATA TGTTAAAAGC ATCTCTCTGA ATTAgAGAGG GAATCTGTcA 154446027
CATCTGTCAC TAATATTTTA GAACAGGCCA CCCCGATCCA TCTTtagTGA 154446077
GTGGAGCATC TCTGCCTgAA AACATCTATA TCCAAATCTt TCTTTCTTTC 154446127
TTTCtttctt tctttttttg atacagcgtc tcgccctatt gcccaggctg 154446177
gagtgcagtg acgtgatctc ggctcactgc aacctctgcc tcccaggttc 154446227
aggc

F4 (Ctr 2)
cagctgttta cattagcaat tacattgaat tagggattat aagtagtcta 154445821
gaggtggctt aaaggatacg ggaggatgtg cttacattat atgcaaatac 154445871
CACACATTTT ATATCAaGGA CTTGAGCATC TGTGGATTTT GGTATCTGCA 154445921
GGGGTGTCCC GGAACCAATC TTCCATGGAT ACCAAGGATG ACtGTGCTCA 154445971
TATTTGTGAT CATATATGTT AAAAGCATCT CTCTGAATTA gAGAGGGAAT 154446021
CTGTCACATC TGTCACTAAT ATTTTAGAAC AGGCCACCCC GATCCATCtT 154446071
TAgTGAGTGG AGCATCTCTG CCTgAAAACA TCTATATCCA AATCTtTCTT 154446121
tCTTTCTTTC TTTCtttctt tttttgatac agcgtctcgc cctattgccc 154446171
aggctggagt gcagtgacgt gatctcggct cactgcaacc tctgcctccc 154446221
aggttcaggc gatt

E4 (Ctr 1)
acagctgttt acattagcaa ttacattgaa ttagggatta taagtagtct 154445820
agaggtggct taaaggatac gggaggatgt gcttacatta tatgcaaata 154445870
CCACACATTT TATATCAAGG ACTTGAGCAT CTGTGGATTT TGGTATCTGC 154445920
AGGGGTGTCC CGgAACCAAT CTTCCATGGA TACCAAGGAT GACTGTGCTC 154445970
ATATTTGTGA TCATATATGT tAAAAGCATC TCTCTGAATT AgAGAGGGAA 154446020
TCTGTCACAT CTGTCACTAA TATTTTAGAA CAGGCCACCC CGATCCATCt 154446070
ttagtGAGTG GAGCATCTCT GCCTgAAAAC ATCTATATCC AAATCTttCT 154446120
TTCTTTCTTT CTTtctttct ttttttgata cagcgtctcg ccctattgcc 154446170
caggctggag tgcagtgacg tgatctcggc tcactgcaac ctctgcctcc 154446220
caggttcagg cga

C4 (Het 2)
tttacattag caattacatt gaattaggga ttataagtag tctagaggtg 154445827
gcttaaagga tacgggagga tgtgcttaca ttatatgcaa ataccacaca 154445877
TTTTATATCA AGGACTTGAG CATCTGTGGA TTTTGGTATC TGCAGGGGTG 154445927
TCCCGGAACC AATCTTCCAT GGATACCAAG GATGACTGTG CTCATATTTG 154445977
TGATCATATA TGTtAAAAGC ATCTCTCTGA ATTAgAGAGG GAATCTGTCa 154446027
CATCTGTCAC TAATATTTTA GAACAGGCCA CCCCGATCCA TCtTTAgTGA 154446077
GTGGAGCATC TCTGCCTgAA AACATCTATA TCCAAATCTT TCTTTCTTTC 154446127
TTTCTTtctt tctttttttg atacagcgtc tcgccctatt gcccaggctg 154446177
gagtgcagtg acgtgatctc ggctcactgc aacctctgcc tcccaggttc 154446227
aggcga

A8 (Het 1)
tagcaattac attgaattag ggattataag tagtctagag gtggcttaaa 154445834
ggatacggga ggatgtgctt acattatatg caaataccac acattttata 154445884
TCAAGGACTT GAGCATCTGT GGATTTTGGT ATCTGCAGGG GTGTCCCGGA 154445934
ACCAATCTTC CATGGATACC AAGGATGACT GTGCTCATAT TTGTGATCAT 154445984
ATATGTTAAA AGCATCTCTC TGAATTAgAG AGGGAATCTG TCaCATCTGT 154446034
CACTAATATT TTAGAACAGG CCACCCCGAT CCATCTTTAg TGAGTgGAGC 154446084
ATCTCTGCCT gAAAACATCT ATATCCAAAT CTTTCTTTCT TTCTTTCttt 154446134
ctttcttttt ttgatacagc gtctcgccct attgcccagg ctggagtgca 154446184
gtgacgtgat ctcggctcac tgcaacctct gcctcccagg ttcaggc

G11 (Het 3)
gctgtttaca ttagcaatta cattgaatta gggattataa gtagtctaga 154445823
ggtggcttaa aggatacggg aggatgtgct tacattatat gcaaatacca 154445873
CACATTTTAT ATCAAGGACT TGAGCATCTG TGGATTTTGG TATCTGCAGG 154445923
GGTGTCCCGG AACCAATCTT CCATGGATAC CAAGGATGAC TGTGCTCATA 154445973
TTTGTGATCA TATATGTTAA AAGCATCTCT CTGAATTAgA GAGGGAATCT 154446023
GTCaCATCTG TCACTAATAT TTTAGAACAG GCCACCCCGA TCCATCtTTA 154446073
gTGAGTgGAG CATCTCTGCC TgAAAACATC TATATCCAAA TCTTTCTTTC 154446123
TTTCtttctt tctttctttt tttgatacag cgtctcgccc tattgcccag 154446173
gctggagtgc agtgacgtga tctcggctca ctgcaacctc tgcctcccag 154446223
gttc

***IGSF9B* (intron)**

E11 (Ctr 3)
Failed

F4 (Ctr 2)
actgggcttt ggtaggatac atagtctctc tggcactgac agcaggactg 133814804
ggaggggacc ctagcatcct ggaccccagc ctgccttccc tctgccctag 133814754
AGAGGGAAGC TGGGtgagcc agtagTCCTG GAAATGCTGc tgggataggg 133814704
atattgctca gcctgaatgg aggtgctCCC GGGTGCGCTg GAGACCCCCC 133814654
ACCcATCCTC TTCTGTTGGC aCTTTTtCAT TcTCTCTTtC ATCTCttcac 133814604
agctctgtat ccatcatgcc ttacttttgt ctcaggagac ctccaaaaga 133814554
atgagagtat tctagggaac tgaggctgct ctcaatgcca agtgc

E4 (Ctr 1)
ggcttctgca cctctggcct cctcattgcc acatgtgaga cacagatggc 133814973
atcctactgg ggcaggggat tagagctgag gaagggtgcc tccaggatgt 133814923
TTGCTCTGAT GGCCAaCCCA TTCCTTGTTC TCaGGCTTCt GGGCCCTGCT 133814873
GCCCTCATCT CCTGGGGTCA CTGGGCTTTG gtagGATACA TAgTCTCTCT 133814823
GGCACTGACa gcagGACTGG GAgGGGACCC TAgCATCCTG gACCCCAGCC 133814773
TGCCTTCCCT CTGCCCTAgA GAGGGAAGCT GGgtgagcca gtagTCCTgG 133814723
AAATGCTGCT GGgATAgGGA TATTGCTCAG CCtGAATGGA gGTGCTCCCg 133814673
GGTGCGCTGG AGACCCCCCA CCCaTCCtCT TCTGTTGGCA CTTTTtCaTT 133814623
CTCTCTTTCA TCTCTtcaca gctctgtatc catcatgcct tacttttgtc 133814573
tcaggagacc tccaaaagaa tgagagtatt ctagggaact gaggctgctc 133814523
tcaatgccaa gtgca

C4 (Het 2)

```
tgtgagacac agatggcatc ctactggggc aggggattag agctgaggaa  133814940
gggtgcctcc aggatgtttg ctctgatggc caacccattc cttgttctca  133814890
GGCTTCTGGG CCCTGCTGCC CTCATCTCCT GGGGTCACTG GGCTTTGGTA  133814840
GgATACATAg TCTCTCTGGC ACTGACAgCA gGACTGGGAG GGGACCCTAg  133814790
CATCCTGgAC CCCAGCCTGC CTTCCCTCTG CCCTagAGAG GGAAGCTGGg  133814740
tgagccagta gtcctggaaa tgctgctggg atagggATAT TGCTCAGcct  133814690
gaatggaggt GCTCCCgGGT GCGCTGGAGA CCCCCCACcc atcctcttct  133814640
gttggcactt tttcattctc tctttcatct cttcacagct ctgtatccat  133814590
catgccttac ttttgtctca ggagacctcc aaaagaat
```

A8 (Het 1)

```
gaggcttctg cacctctggc ctcctcattg ccacatgtga gacacagatg  133814975
gcatcctact ggggcagggg attagagctg aggaagggtg cctccaggat  133814925
GTTTGCTCtG ATGGCCaACC CATTCCTTGT tctcaGGCTT CTGGGCCCTG  133814875
CTGCCCTCAT CTCCTGGGGT CACTGGGCTT TGGTAGgATA CATAgTCTCT  133814825
CTGGCACTGA CagcagGACT GGGAGGGGAC CCTAgCATCC TGgACCCcAG  133814775
CCTGCCTTCC CTCTGCCCTA gAGAGGGAAG CTGGgtgagc cagtagTCCT  133814725
GgAAATGCTG CTGGgATAgG GATATTGCTC AGCCtgaatG GAgGtGCTCC  133814675
CgGGTGCGCT GGAGACCCCC CACCCaTCCT CTTctGTTGG CaCTTTTtCa  133814625
TTCTCTCTTt CATCTCttca cagctctgta tccatcatgc cttacttttg  133814575
tctcaggaga cctccaaaag aatgagagta ttctagggaa ctgaggctgc  133814525
tctcaatgcc aagtgc
```

G11 (Het 3)

```
ggcttctgca cctctggcct cctcattgcc acatgtgaga cacagatggc  133814973
atcctactgg ggcaggggat tagagctgag gaagggtgcc tccaggatgt  133814923
TTGCTCTGAT GGCCaACCCA TTCCTTGTTC TCAGGCTTCT GGGCCCTGCT  133814873
GCCCTCATCT CCTGGGGTCA CTGGGCTTTG GTAGGATACA TAGTCTCTCT  133814823
GGCACTGACA GCAGGACTGG GAGGGGACCC TAGCATCCTG GACCCCAGCC  133814773
TGCCTTCCCT CTGCCCTAGA GAGGGAAGCT GGGTGAGCCA GTAGTCCTGG  133814723
AAATGCTGCT GGGATAGGGA TATTGCTCAG CCTGAATGGA GGTGCTCCCG  133814673
GGTGCGCTGG AGACCCCCCA CCCATCCTCT TCTGTTGGCA CTTTTTCATT  133814623
CTCTCTTTCA TCTCTTCACA GCTCTGTATC CATCATGCCT tacttttgtc  133814573
tcaggagacc tccaaaagaa tgagagtatt ctagggaact gaggctgctc  133814523
tcaatgccaa gtgcattgag ttgccttgca ggctggggca
```

TSPAN11

E11 (Ctr 3)

```
ccccgggcac gctggacggc acgaggccct cagaggccca gagctctctc  31080020
ccaaggcaca agtgcagaaa ggcagactca tgcgcatagt tatggacacc  31079970
GTGTTGCGTG TccCATGCAG ACACCAATAc CTGCTCACAT GCCCCGCGAG  31079920
TGTcCCAGGC ATTCCTGCCT CCTCCACTCC tcatccccag ccggcttccc  31079870
tctgcccgcc cgggggAAGC TCGGgacaGT CTAGCTCGGG ACTGCcGGCG  31079820
gggcgggcag cgGAGGtGGA GGCGcCTCTC CtGGACCCCC AgCCCCCTCC  31079770
CGCggcGCCC CCACTCCTCG GGCGCGCTTC TGCACTTACC CTGCCAgGGG  31079720
CTCCggaagc ggcgaaggga gctgcgccta gagagactga gagcggcggc  31079670
tcccggggcc gcccagccgc ccaccgcccg cagccagcgc tcctccctcc  31079620
cagc
```

F4 (Ctr 2)

```
ccccgggcac gctggacggc acgaggccct cagaggccca gagctctctc  31080020
ccaaggcaca agtgcagaaa ggcagactca tgcgcatagt tatggacacc  31079970
GTGTTGCGTG TcCCATGCAG ACACCAATAC cTGCTCaCAT GCCCCGCgAG  31079920
TGTCcCAGGC ATTCCTGCCT CCTCCACTCC TCATCCCCAG CCGGCTTCCC  31079870
TCTGCCCGcC cgGGGGAaGC TCGGGAcagT CTAGCTCGGG ACTGCCGgcg  31079820
gggcgggcag cggaggtgGA GGCGCCTCTC CTGgACCCCC AGCCCCCTCC  31079770
CGCGGCGCCC CCACTCCTCG GGCGCGCTTC TGCACTTACC CTGCCAGGGG  31079720
```

CTCCggaagc ggcgaaggga gctgcgccta gagagactga gagcggcggc 31079670
tcccggggcc gcccagccgc ccaccgcccg cagccagcgc tcctccctcc 31079620
cagc

E4 (Ctr 1)
tgccccgggc acgctggacg gcacgaggcc ctcagaggcc cagagctctc 31080022
tcccaaggca caagtgcaga aaggcagact catgcgcata gttatggaca 31079972
CCGTGTTGCG TGTCCCATGC AGACACCAAT AcCTGCTCAC ATGCCCCGCG 31079922
AGTGTCCCAG GCATTCCTGC CTCCTCCACT CCTCATCCCC AGCCGGCTTC 31079872
CCTCTGCCCG CcCGGGGGAA GCTCGGGACA GTCtaGCTCG GGACTGCCGG 31079822
CGgggcgggc agcgGAGGtG GAGGCGCCTC TCCTGgACCC cCAgCCCCCT 31079772
CCCGCGGCGC CCCCACTCCT CGGGCGCGCT TCTGCACTTA CCCTGCCAGG 31079722
GGCTCCggaa gcggcgaagg gagctgcgcc tagagagact gagagcggcg 31079672
gctcccgggg ccgcccagcc gcccaccgcc cgcagccagc gctcctccct 31079622
cccagc

C4 (Het 2)
gccccgggca cgctggacgg cacgaggccc tcagaggccc agagctctct 31080021
cccaaggcac aagtgcagaa aggcagactc atgcgcatag ttatggacac 31079971
CGTGTTGCGT GTcCCATGCA GACACCAATA cCTGCTCACA TGCCCCGCGA 31079921
GTGtcCCAGG CATTCCTGCC TCCTCCACTC CTCATCCCCA GCCgGCTTCC 31079871
CTCTGCCCGc cCGGGGGAAG CTCGGGAcag tCtaGCTCGG GACTGCCGgc 31079821
ggggcgggca gcgGAGGtGG AGGCGCCTCT CCTGgACCCc CAgCCCCCTC 31079771
CCGCGGCGCC CCCACTCCTC GGGCGCGCTT CTGCACTTAC CCTGCCAGGG 31079721
GCTCCGGAAg CGGCgAAGGg AGCTGCgCCT Agagagactg agagcggcgg 31079671
ctcccggggc cgcccagccg cccaccgccc gcagccagcg ctcctccctc 31079621
ccagcccggg aaggtcagcg tgtgggcagc c

A8 (Het 1)
gccccgggca cgctggacgg cacgaggccc tcagaggccc agagctctct 31080021
cccaaggcac aagtgcagaa aggcagactc atgcgcatag ttatggacac 31079971
CGTGTTGCGT GTcCCATGCA GACACCAATA CCTGCTCACA TGCCCCGCGA 31079921
GTGTcCCAGG CATTCCTGCC TCCTCCACTC CTCATCCCCA GCCGGCTTCC 31079871
CTCTGCCCGc ccgggggaaG CTCGGGAcag tctagctCGG GACTGCCGgc 31079821
ggggcgggca gcggaggtgg aggcgcCTCT CCTGgACCCc CAGCCCCCTC 31079771
CCGCGGCGCC CCCACTCCTc GGGCGCGCTT CTGCACTTAC CCTGCCAGGG 31079721
GCTCCGGAAg CGgcgaaggg agctgcgcct agagagactg agagcggcgg 31079671
ctcccggggc cgcccagccg cccaccgccc gcagccagcg ctcctccctc 31079621
ccagcccggg aaggtcagcg tgtgggcagc ccggcccgcg cccctgcgcc 31079571

G11 (Het 3)
ccccgggcac gctggacggc acgaggccct cagaggccca gagctctctc 31080020
ccaaggcaca agtgcagaaa ggcagactca tgcgcatagt tatggacacc 31079970
GTGTTGCGTG TcCCATGCAG ACACCAATAc CTGCTCACAT GCCCCGCGAG 31079920
TGTCcCAGGC ATTCCTGCCT CCTCCACTCC TCATCCCCAG CCGGCTTCCC 31079870
TCTGCCCgCc CGGGGGAAGC TCGGGAcagt CtaGCTCGGG ACTGCCGGCG 31079820
GGgcgggcag cgGAGGtGGA GGCGCCTCTC CTGgaCCCcC AGCCCCCTCC 31079770
CGCGGCGCCC CCACTCCTCG GGCGCGCTTC TGCACTTACC CTGCCAGGGG 31079720
CTCCGGAAgC GGCgAAGGga gctgcgccta gagagactga gagcggcggc 31079670
tcccggggcc gcccagccgc ccaccgcccg cagccagcgc tcctccctcc 31079620
cagcccggga aggtcagc

Appendix 9: Allelic Imbalance table, raw data

Samples	rs2981578	rs1047100	rs1047100 cDNA		rs1047100 gDNA		cDNA-gDNA Ct		Difference	Absolute difference
			Ctvalue Allele 1	Ctvalue Allele 2	Ctvalue Allele 1	Ctvalue Allele 2	Ctvalue Allele 1	Ctvalue Allele 2		
15	C/T	C/T	30.24	29.89	28.59	27.68	1.65	2.21	-0.56	0.56
36	C/T	C/T	26.18	25.44	29.31	28.08	-3.13	-2.64	-0.49	0.49
39	C/T	C/T	28.07	27.31	27.89	26.67	0.18	0.65	-0.47	0.47
59	C/T	C/T	31.84	29.62	27.56	25.26	4.28	4.35	-0.07	0.07
62	C/T	C/T	29.24	29.35	25.71	25.07	3.53	4.28	-0.75	0.75
									total	**0.41**
Ctr	rs2981578	rs1047100								
12	C/C	C/T	32.90	34.33	30.92	32.56	1.98	1.77	0.21	0.21
26	C/C	C/T	30.08	30.64	31.30	31.28	-1.22	-0.64	-0.58	0.58
27	C/C	C/T	29.40	30.80	31.26	31.56	-1.86	-0.76	-1.10	1.1
29	C/C	C/T	27.65	28.79	30.68	30.48	-3.03	-1.69	-1.34	1.34
44	C/C	C/T	31.42	32.98	30.99	31.35	0.43	1.63	-1.20	1.20
61	C/C	C/T	31.53	31.59	27.00	27.32	4.53	4.27	0.26	0.26
65	C/C	C/T	35.83	36.13	27.93	28.13	7.90	8.00	-0.10	0.1
28	T/T	C/T	30.05	32.39	29.07	29.30	0.98	3.09	-2.11	2.11
38	T/T	C/T	30.28	31.69	30.00	29.95	0.28	1.74	-1.46	1.46
70	T/T	C/T	33.72	35.71	29.25	29.46	4.47	6.25	-1.78	1.78
									total	**1.014**
53	T/T	C/T	30.84	N/A	26.64	28.23	4.20	N/A	N/A	N/A
6	C/C?	C/T	35.08	35.69	29.73	33.15	5.35	2.54	2.81	2.81

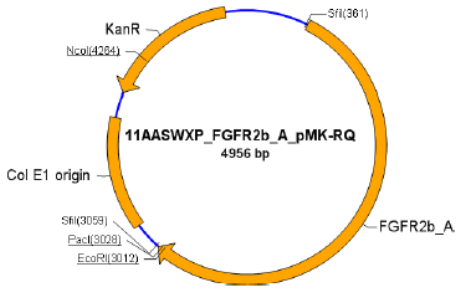
Appendix 10: FGFR2b-GFP construct

The construct was assembled from two Gene Art plasmids Gene ART project for GFP-tagged FGFR2b construct. The 2 fragment were synthesised by Invitrogen and provided as two vectors.

Plasmid DNA Description:

The synthetic gene FGFR2b_A was assembled from synthetic oligonucleotides and/or PCR products. The fragment was cloned into pMK-RQ (kanR) using SfiI and SfiI cloning sites. The plasmid DNA was purified from transformed bacteria and concentration determined by UV spectroscopy. The final construct was verified by sequencing. The sequence congruence within the used restriction sites was 100%. See the accompanying data sheets for sequences and find the original ABI trace files as well as the assembled sequences electronically on disk.
5 µg of the plasmid preparation were lyophilized for shipping.

Plasmid Map:



life technologies™

Quality Assurance Documentation: 11AASWXP

Ref. No.: 1165073

Designation:	E.coli K12 (dam+ dcm+ tonA rec-)
Gene name:	FGFR2b_A
Gene size:	2678 bp
Vector backbone:	pMK-RQ (kanR)
Cloning sites:	SfiI / SfiI
Quantity:	~5 µg Plasmid DNA
Note:	Please dissolve lyophilized DNA in 50 µl distilled water or 10 mM Tris-HCl (pH 8.0). We recommend sequence verification after each transformation step.
Date:	5 October 2011

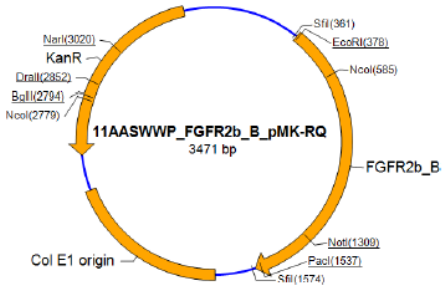
Martina Siefken
Quality control

GeneArt AG **www.lifetechnologies.com** **GeneArtSupport@lifetech.com**

Plasmid DNA Description:

The synthetic gene FGFR2b_B was assembled from synthetic oligonucleotides and/or PCR products. The fragment was cloned into pMK-RQ (kanR) using SfiI and SfiI cloning sites. The plasmid DNA was purified from transformed bacteria and concentration determined by UV spectroscopy. The final construct was verified by sequencing. The sequence congruence within the used restriction sites was 100%. See the accompanying data sheets for sequences and find the original ABI trace files as well as the assembled sequences electronically on disk.
5 µg of the plasmid preparation were lyophilized for shipping.

Plasmid Map:



life technologies™

Quality Assurance Documentation: 11AASWWP

Ref. No.: 1165072

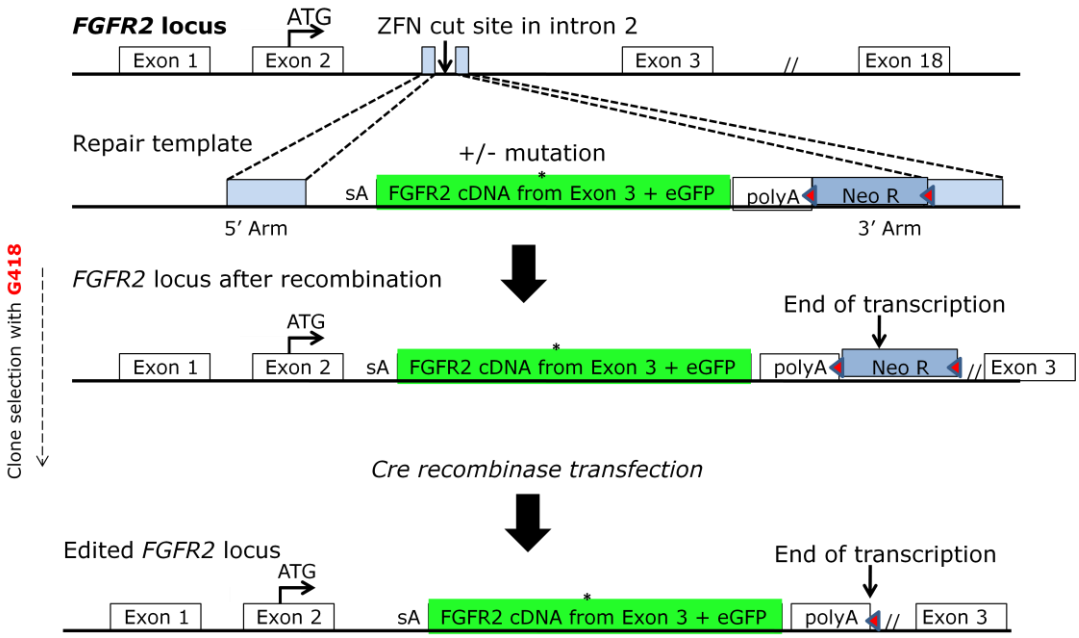
Designation:	E.coli K12 (dam+ dcm+ tonA rec-)
Gene name:	FGFR2b_B
Gene size:	1193 bp
Vector backbone:	pMK-RQ (kanR)
Cloning sites:	SfiI / SfiI
Quantity:	~5 µg Plasmid DNA
Note:	Please dissolve lyophilized DNA in 50 µl distilled water or 10 mM Tris-HCl (pH 8.0). We recommend sequence verification after each transformation step.
Date:	6 October 2011

Martina Siefken
Quality control

GeneArt AG **www.lifetechnologies.com** **GeneArtSupport@lifetech.com**

Appendix 11: FGFR2b-GFP/neo construct targeted integration

Diagram showing the structure of the FGFR2b-GFP construct and the insertion locus in the second intron of the FGFR2 gene.



FGFR2b-GFP construct

The coding region of the GFP-tagged FGFR2b construct was synthesised by Invitrogen (GeneART project) as two distinct inserts cloned into PMK-RQ vector. The vector FGFR2b_A was opened by a double digestion using *EcoRI* and *PacI* and purified. The insert FGFR2b_B was excised from the vector using double digestion with *EcoRI* and *PacI*. This generated compatible ends for cloning the second insert into the first vector (1:3 ratio) in the following ligation reaction, carried out overnight at 4°C:

T4 Ligase.....1 µl

10X Ligase buffer (with ATP).....1 µl

Vector FGFR2b_A.....2 µl

Insert FGFR2b_B.....6 µl

Insert A+B was lifted out of PMK-RQ using *SpeI* restriction enzyme digestion and cloned in MCF7 repair template with a unique *SpeI* site in the middle of the 2kb homology region, previously created by site-directed mutagenesis (SDM). In a later

stage, a neomycin resistance cassette was added (from PGKneolox2DTA.2 vector, Addgene) after the poly(A) tail of the cDNA using a *BamHI* restriction site.

Transfection in T47D cells

ZFN-edited T47D cells were seeded in 96 well plates at a concentration of 4 cells/ml, with 100 µl of complete medium containing 600 µg/ml of G418 antibiotic (Sigma) in each well. After the single cell colonies reached 50 to 100 cells, they were detached by trypsination and transferred to a new 96 well plate and duplicated. A duplicate of each plate containing monoclonal MCF10A and MCF7 cells was made 7 days later and sent to the Genome Centre (Barts Cancer Institute, QMUL) for gDNA extraction and Taqman SNP genotyping assay for rs2981578, once the cells were confluent. The edited T47D cells were screened for GFP expression, as observed under UV light microscopy.

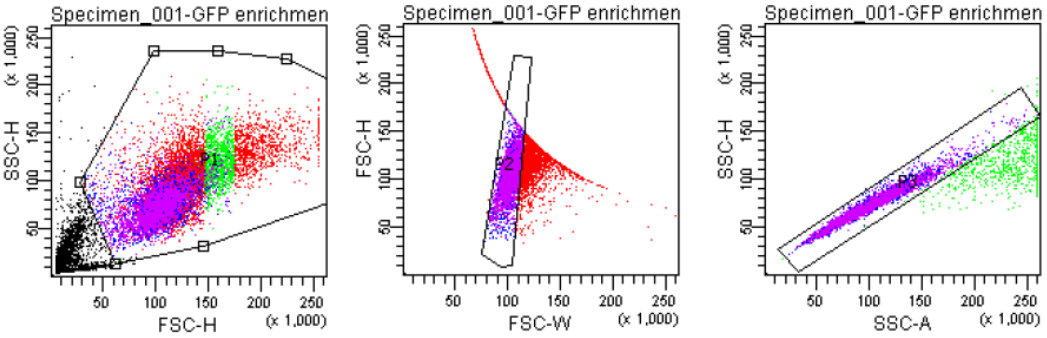
Appendix 12: Culture media used for each cell lines

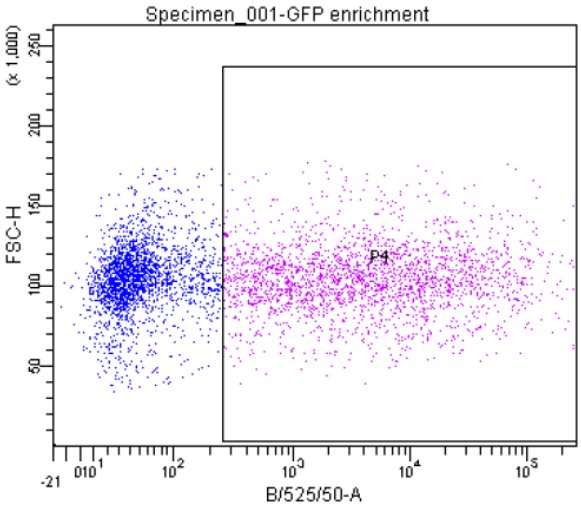
Cell line	Culture media
AU561	RPMI, 10% FBS
BT20	DMEM, 10%FBS
BT474	RPMI, 10% FBS
Cal51	RPMI, 10% FBS
H3396	DMEM, 10%FBS
HFF2	DMEM, 10%FBS
MCF10A	DMEM, F12*
MCF7	DMEM, 10%FBS
MDA-MB-231	DMEM, 10%FBS
MDA-MB-453	DMEM, 10%FBS
MDA-MB-468	L15, 10% FBS
SKBR3	RPMI, 10% FBS
SUM159	Ham's F12, 5%FBS, IH[#]
T47D	RPMI, 10% FBS
ZR-75-1	RPMI, 10% FBS
β4-1089	DMEM, 10%FBS

* DMEM/F12 media supplemented with 5% horse serum, 20 ng/ml EGF, 100 ng/ml cholera enterotoxin, 0.01 mg/ml insulin and 500 ng/ml hydrocortisone.

[#] Ham's F12 media supplemented with 5% FBS, insuling (0.01 mg/ml) and hydrocortisone (500 ng/ml)

Appendix 13: FACS GFP sorting

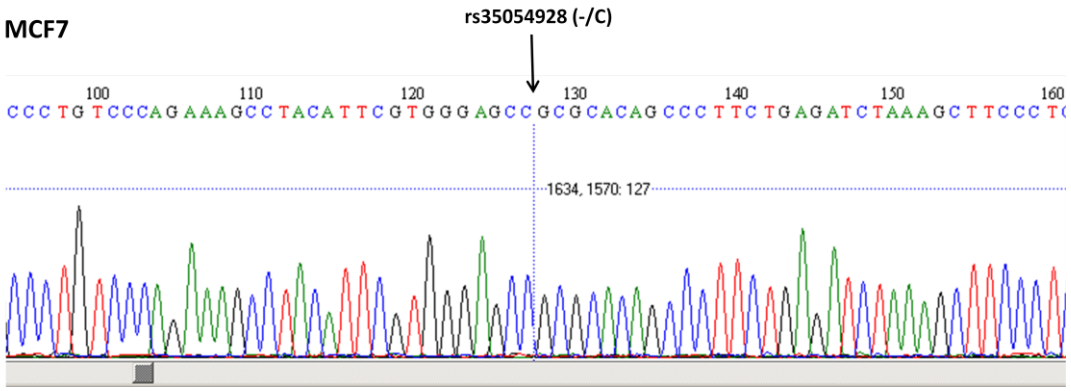




Tube: GFP enrichment

Population	#Events	%Parent	%Total
All Events	15,046	####	100.0
P1	12,036	80.0	80.0
P2	5,912	49.1	39.3
P3	4,760	80.5	31.6
P4	2,449	51.4	16.3

Appendix 14: rs35054928 genotype in MCF7 cells



Appendix 15: Previous publication

Fibroblast growth factor 22 is not essential for skin development and repair but plays a role in tumorigenesis.

Jarosz M, Robbez-Masson L, Chioni AM, Cross B, Rosewell I, Grose R.

PLoS One. 2012;7(6):e39436. doi: 10.1371/journal.pone.0039436.

CHAPTER 8

REFERENCES

8. References

1000 Genomes Project, C. (2008-2012) 1000 Genomes project. *http://www.1000genomes.org/about,* 06.06.12.

1000Genomes (2011) The 1000 Genomes project, minor allele frequencies in different populations. *http://www.1000genomes.org/,* 20.09.12.

Abelson, J. F., Kwan, K. Y., O'Roak, B. J., Baek, D. Y., Stillman, A. A., Morgan, T. M., Mathews, C. A., Pauls, D. L., Rasin, M. R., Gunel, M., Davis, N. R., Ercan-Sencicek, A. G., Guez, D. H., Spertus, J. A., Leckman, J. F., Dure, L. S. t., Kurlan, R., Singer, H. S., Gilbert, D. L., Farhi, A., Louvi, A., Lifton, R. P., Sestan, N. & State, M. W. (2005) Sequence variants in SLITRK1 are associated with Tourette's syndrome. *Science,* 310**,** 317-20.

Adelaide, J., Finetti, P., Bekhouche, I., Repellini, L., Geneix, J., Sircoulomb, F., Charafe-Jauffret, E., Cervera, N., Desplans, J., Parzy, D., Schoenmakers, E., Viens, P., Jacquemier, J., Birnbaum, D., Bertucci, F. & Chaffanet, M. (2007) Integrated profiling of basal and luminal breast cancers. *Cancer Res,* 67**,** 11565-75.

Ahmed, Z., George, R., Lin, C. C., Suen, K. M., Levitt, J. A., Suhling, K. & Ladbury, J. E. (2010) Direct binding of Grb2 SH3 domain to FGFR2 regulates SHP2 function. *Cell Signal,* 22**,** 23-33.

Alimonti, A., Carracedo, A., Clohessy, J. G., Trotman, L. C., Nardella, C., Egia, A., Salmena, L., Sampieri, K., Haveman, W. J., Brogi, E., Richardson, A. L., Zhang, J. & Pandolfi, P. P. (2010) Subtle variations in Pten dose determine cancer susceptibility. *Nat Genet,* 42**,** 454-8.

Andre, F., Job, B., Dessen, P., Tordai, A., Michiels, S., Liedtke, C., Richon, C., Yan, K., Wang, B., Vassal, G., Delaloge, S., Hortobagyi, G. N., Symmans, W. F., Lazar, V. & Pusztai, L. (2009) Molecular characterization of breast cancer with high-resolution oligonucleotide comparative genomic hybridization array. *Clin Cancer Res,* 15**,** 441-51.

Antoniou, A., Pharoah, P. D., Narod, S., Risch, H. A., Eyfjord, J. E., Hopper, J. L., Loman, N., Olsson, H., Johannsson, O., Borg, A., Pasini, B., Radice, P., Manoukian, S., Eccles, D. M., Tang, N., Olah, E., Anton-Culver, H., Warner, E., Lubinski, J., Gronwald, J., Gorski, B., Tulinius, H., Thorlacius, S., Eerola, H., Nevanlinna, H., Syrjakoski, K., Kallioniemi, O. P., Thompson, D., Evans, C., Peto, J., Lalloo, F., Evans, D. G. & Easton, D. F. (2003) Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case Series unselected for family history: a combined analysis of 22 studies. *Am J Hum Genet,* 72**,** 1117-30.

Antoniou, A. C. & Easton, D. F. (2006) Models of genetic susceptibility to breast cancer. *Oncogene,* 25**,** 5898-905.

Antoniou, A. C., Spurdle, A. B., Sinilnikova, O. M., Healey, S., Pooley, K. A., Schmutzler, R. K., Versmold, B., Engel, C., Meindl, A., Arnold, N., Hofmann, W., Sutter, C., Niederacher, D., Deissler, H., Caldes, T., Kampjarvi, K., Nevanlinna, H., Simard, J., Beesley, J., Chen, X., Neuhausen, S. L., Rebbeck, T. R., Wagner, T., Lynch, H. T., Isaacs, C., Weitzel, J., Ganz, P. A., Daly, M. B., Tomlinson, G., Olopade, O. I., Blum, J. L., Couch, F. J., Peterlongo, P., Manoukian, S., Barile, M., Radice, P., Szabo, C. I., Pereira, L. H., Greene, M. H., Rennert, G., Lejbkowicz, F., Barnett-Griness, O.,

Andrulis, I. L., Ozcelik, H., Gerdes, A. M., Caligo, M. A., Laitman, Y., Kaufman, B., Milgrom, R., Friedman, E., Domchek, S. M., Nathanson, K. L., Osorio, A., Llort, G., Milne, R. L., Benitez, J., Hamann, U., Hogervorst, F. B., Manders, P., Ligtenberg, M. J., van den Ouweland, A. M., Peock, S., Cook, M., Platte, R., Evans, D. G., Eeles, R., Pichert, G., Chu, C., Eccles, D., Davidson, R., Douglas, F., Godwin, A. K., Barjhoux, L., Mazoyer, S., Sobol, H., Bourdon, V., Eisinger, F., Chompret, A., Capoulade, C., Bressac-de Paillerets, B., Lenoir, G. M., Gauthier-Villars, M., Houdayer, C., Stoppa-Lyonnet, D., Chenevix-Trench, G. & Easton, D. F. (2008) Common breast cancer-predisposition alleles are associated with breast cancer risk in BRCA1 and BRCA2 mutation carriers. *Am J Hum Genet,* 82**,** 937-48.

Badve, S., Dabbs, D. J., Schnitt, S. J., Baehner, F. L., Decker, T., Eusebi, V., Fox, S. B., Ichihara, S., Jacquemier, J., Lakhani, S. R., Palacios, J., Rakha, E. A., Richardson, A. L., Schmitt, F. C., Tan, P. H., Tse, G. M., Weigelt, B., Ellis, I. O. & Reis-Filho, J. S. (2011) Basal-like and triple-negative breast cancers: a critical review with an emphasis on the implications for pathologists and oncologists. *Mod Pathol,* 24**,** 157-67.

Bane, A. L., Pinnaduwage, D., Colby, S., Reedijk, M., Egan, S. E., Bull, S. B., O'Malley, F. P. & Andrulis, I. L. (2009) Expression profiling of familial breast cancers demonstrates higher expression of FGFR2 in BRCA2-associated tumors. *Breast Cancer Res Treat,* 117**,** 183-91.

Barnholtz-Sloan, J. S., Raska, P., Rebbeck, T. R. & Millikan, R. C. (2011) Replication of GWAS "Hits" by Race for Breast and Prostate Cancers in European Americans and African Americans. *Front Genet,* 2**,** 37.

Barnholtz-Sloan, J. S., Shetty, P. B., Guan, X., Nyante, S. J., Luo, J., Brennan, D. J. & Millikan, R. C. (2010) FGFR2 and other loci identified in genome-wide association studies are associated with breast cancer in African-American and younger women. *Carcinogenesis,* 31**,** 1417-23.

Bauer, K. R., Brown, M., Cress, R. D., Parise, C. A. & Caggiano, V. (2007) Descriptive analysis of estrogen receptor (ER)-negative, progesterone receptor (PR)-negative, and HER2-negative invasive breast cancer, the so-called triple-negative phenotype: a population-based study from the California cancer Registry. *Cancer,* 109**,** 1721-8.

Bernstein, B. E., Birney, E., Dunham, I., Green, E. D., Gunter, C. & Snyder, M. (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature,* 489**,** 57-74.

Bevier, M., Sundquist, K. & Hemminki, K. (2012) Risk of breast cancer in families of multiple affected women and men. *Breast Cancer Res Treat,* 132**,** 723-8.

Bibikova, M., Carroll, D., Segal, D. J., Trautman, J. K., Smith, J., Kim, Y. G. & Chandrasegaran, S. (2001) Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. *Mol Cell Biol,* 21**,** 289-97.

Bibikova, M., Golic, M., Golic, K. G. & Carroll, D. (2002) Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases. *Genetics,* 161**,** 1169-75.

Birling, M. C., Gofflot, F. & Warot, X. (2009) Site-specific recombinases for manipulation of the mouse genome. *Methods Mol Biol,* 561**,** 245-63.

Bond, G. L., Hirshfield, K. M., Kirchhoff, T., Alexe, G., Bond, E. E., Robins, H., Bartel, F., Taubert, H., Wuerl, P., Hait, W., Toppmeyer, D., Offit, K. & Levine, A. J. (2006) MDM2 SNP309 accelerates tumor formation in a gender-specific and hormone-dependent manner. *Cancer Res,* 66**,** 5104-10.

Bond, G. L., Hu, W., Bond, E. E., Robins, H., Lutzker, S. G., Arva, N. C., Bargonetti, J., Bartel, F., Taubert, H., Wuerl, P., Onel, K., Yip, L., Hwang, S. J., Strong, L. C., Lozano, G. & Levine, A. J. (2004) A single nucleotide polymorphism in the MDM2 promoter attenuates the p53 tumor suppressor pathway and accelerates tumor formation in humans. *Cell,* 119**,** 591-602.

Bookout, A. L. & Mangelsdorf, D. J. (2003) Quantitative real-time PCR protocol for analysis of nuclear receptor signaling pathways. *Nucl Recept Signal,* 1**,** e012.

Bowen, R. L., Duffy, S. W., Ryan, D. A., Hart, I. R. & Jones, J. L. (2008) Early onset of breast cancer in a group of British black women. *Br J Cancer,* 98**,** 277-81.

Brem, R. B., Yvert, G., Clinton, R. & Kruglyak, L. (2002) Genetic dissection of transcriptional regulation in budding yeast. *Science,* 296**,** 752-5.

Brookes, A. J. (1999) The essence of SNPs. *Gene,* 234**,** 177-86.

Bryson (2012) Range of Ductal Carcinoma in situ (DCIS). *http://www.breastcancer.org/pictures/types/dcis/dcis_range,* 02.11.12.

Byron, S. A., Gartside, M. G., Wellens, C. L., Mallon, M. A., Keenan, J. B., Powell, M. A., Goodfellow, P. J. & Pollock, P. M. (2008) Inhibition of activated fibroblast growth factor receptor 2 in endometrial cancer cells induces cell death despite PTEN abrogation. *Cancer Res,* 68**,** 6902-7.

Calle, E. E. & Kaaks, R. (2004) Overweight, obesity and cancer: epidemiological evidence and proposed mechanisms. *Nat Rev Cancer,* 4**,** 579-91.

Carroll, D. (2008) Progress and prospects: zinc-finger nucleases as gene therapy agents. *Gene Ther,* 15**,** 1463-8.

Carroll, J. S., Liu, X. S., Brodsky, A. S., Li, W., Meyer, C. A., Szary, A. J., Eeckhoute, J., Shao, W., Hestermann, E. V., Geistlinger, T. R., Fox, E. A., Silver, P. A. & Brown, M. (2005) Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell,* 122**,** 33-43.

Carroll, J. S., Lynch, D. K., Swarbrick, A., Renoir, J. M., Sarcevic, B., Daly, R. J., Musgrove, E. A. & Sutherland, R. L. (2003) p27(Kip1) induces quiescence and growth factor insensitivity in tamoxifen-treated breast cancer cells. *Cancer Res,* 63**,** 4322-6.

CCLE (2012) Cancer cell line encyclopedia from the Broad Institute and the Novartis Institutes for Biomedical Research. *http://www.broadinstitute.org/ccle/search/searchResult,* 04.03.12.

Chan, M., Ji, S. M., Liaw, C. S., Yap, Y. S., Law, H. Y., Yoon, C. S., Wong, C. Y., Yong, W. S., Wong, N. S., Ng, R., Ong, K. W., Madhukumar, P., Oey, C. L., Tan, P. H., Li, H. H., Ang, P., Ho, G. H. & Lee, A. S. (2012) Association of common genetic variants with breast cancer risk and clinicopathological characteristics in a Chinese population. *Breast Cancer Res Treat,* 136**,** 209-20.

Chang, C. J. & Bouhassira, E. E. (2012) Zinc-finger nuclease-mediated correction of alpha-thalassemia in iPS cells. *Blood,* 120**,** 3906-14.

Charles Knight, J. (2005) HaploChIP: an in vivo assay. *Methods Mol Biol,* 311**,** 49-60.

Chen, K. & Rajewsky, N. (2007) The evolution of gene regulation by transcription factors and microRNAs. *Nat Rev Genet,* 8**,** 93-103.

Chesler, E. J., Lu, L., Shou, S., Qu, Y., Gu, J., Wang, J., Hsu, H. C., Mountz, J. D., Baldwin, N. E., Langston, M. A., Threadgill, D. W., Manly, K. F. & Williams, R. W. (2005) Complex trait analysis of gene expression uncovers polygenic and pleiotropic networks that modulate nervous system function. *Nat Genet,* 37**,** 233-42.

Cheung, V. G., Nayak, R. R., Wang, I. X., Elwyn, S., Cousins, S. M., Morley, M. & Spielman, R. S. (2010) Polymorphic cis- and trans-regulation of human gene expression. *PLoS Biol,* 8.

Chioni, A. M. & Grose, R. (2009) Negative regulation of fibroblast growth factor 10 (FGF-10) by polyoma enhancer activator 3 (PEA3). *Eur J Cell Biol,* 88**,** 371-84.

Chioni, A. M., Shao, D., Grose, R. & Djamgoz, M. B. (2010) Protein kinase A and regulation of neonatal Nav1.5 expression in human breast cancer cells: activity-dependent positive feedback and cellular migration. *Int J Biochem Cell Biol,* 42**,** 346-58.

Cimoli, G., Malacarne, D., Ponassi, R., Valenti, M., Alberti, S. & Parodi, S. (2004) Meta-analysis of the role of p53 status in isogenic systems tested for sensitivity to cytotoxic antineoplastic drugs. *Biochim Biophys Acta,* 1705**,** 103-20.

Cirillo, L. A., Lin, F. R., Cuesta, I., Friedman, D., Jarnik, M. & Zaret, K. S. (2002) Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol Cell,* 9**,** 279-89.

Ciruna, B. & Rossant, J. (2001) FGF signaling regulates mesoderm cell fate specification and morphogenetic movement at the primitive streak. *Dev Cell,* 1**,** 37-49.

Coleman-Krnacik, S. & Rosen, J. M. (1994) Differential temporal and spatial gene expression of fibroblast growth factor family members during mouse mammary gland development. *Mol Endocrinol,* 8**,** 218-29.

Collins, F. S., Brooks, L. D. & Chakravarti, A. (1998) A DNA polymorphism discovery resource for research on human genetic variation. *Genome Res,* 8**,** 1229-31.

Conner, A. J. & Jacobs, J. M. (1999) Genetic engineering of crops as potential source of genetic hazard in the human diet. *Mutat Res,* 443**,** 223-34.

Consortium, E. P., Dunham, I., Kundaje, A., Aldred, S. F., Collins, P. J., Davis, C. A., Doyle, F., Epstein, C. B., Frietze, S., Harrow, J., Kaul, R., Khatun, J., Lajoie, B. R., Landt, S. G., Lee, B. K., Pauli, F., Rosenbloom, K. R., Sabo, P., Safi, A., Sanyal, A., Shoresh, N., Simon, J. M., Song, L., Trinklein, N. D., Altshuler, R. C., Birney, E., Brown, J. B., Cheng, C., Djebali, S., Dong, X., Dunham, I., Ernst, J., Furey, T. S., Gerstein, M., Giardine, B., Greven, M., Hardison, R. C., Harris, R. S., Herrero, J., Hoffman, M. M., Iyer, S., Kelllis, M., Khatun, J., Kheradpour, P., Kundaje, A., Lassman, T., Li, Q., Lin, X., Marinov, G. K., Merkel, A., Mortazavi, A., Parker, S. C., Reddy, T. E., Rozowsky, J., Schlesinger, F., Thurman, R. E., Wang, J., Ward, L. D., Whitfield, T. W., Wilder, S. P., Wu, W., Xi, H. S., Yip, K. Y., Zhuang, J., Bernstein, B. E., Birney, E., Dunham, I.,

Green, E. D., Gunter, C., Snyder, M., Pazin, M. J., Lowdon, R. F., Dillon, L. A., Adams, L. B., Kelly, C. J., Zhang, J., Wexler, J. R., Green, E. D., Good, P. J., Feingold, E. A., Bernstein, B. E., Birney, E., Crawford, G. E., Dekker, J., Elinitski, L., Farnham, P. J., Gerstein, M., Giddings, M. C., Gingeras, T. R., Green, E. D., Guigo, R., Hardison, R. C., Hubbard, T. J., Kellis, M., Kent, W. J., Lieb, J. D., Margulies, E. H., Myers, R. M., Snyder, M., Starnatoyannopoulos, J. A*., et al.* (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature,* 489**,** 57-74.

Copeland, R. A., Ji, H., Halfpenny, A. J., Williams, R. W., Thompson, K. C., Herber, W. K., Thomas, K. A., Bruner, M. W., Ryan, J. A., Marquis-Omer, D. & et al. (1991) The structure of human acidic fibroblast growth factor and its interaction with heparin. *Arch Biochem Biophys,* 289**,** 53-61.

Corson, L. B., Yamanaka, Y., Lai, K. M. & Rossant, J. (2003) Spatial and temporal patterns of ERK signaling during mouse embryogenesis. *Development,* 130**,** 4527-37.

Coulier, F., Pontarotti, P., Roubin, R., Hartung, H., Goldfarb, M. & Birnbaum, D. (1997) Of worms and men: an evolutionary perspective on the fibroblast growth factor (FGF) and FGF receptor families. *J Mol Evol,* 44**,** 43-56.

Coussens, L., Yang-Feng, T. L., Liao, Y. C., Chen, E., Gray, A., McGrath, J., Seeburg, P. H., Libermann, T. A., Schlessinger, J., Francke, U. & et al. (1985) Tyrosine kinase receptor with extensive homology to EGF receptor shares chromosomal location with neu oncogene. *Science,* 230**,** 1132-9.

Cowper-Sal lari, R., Zhang, X., Wright, J. B., Bailey, S. D., Cole, M. D., Eeckhoute, J., Moore, J. H. & Lupien, M. (2012) Breast cancer risk-associated SNPs modulate the affinity of chromatin for FOXA1 and alter gene expression. *Nat Genet,* 44**,** 1191-8.

Cradick, T. J., Ambrosini, G., Iseli, C., Bucher, P. & McCaffrey, A. P. (2011) ZFN-site searches genomes for zinc finger nuclease target sites and off-target sites. *BMC Bioinformatics,* 12**,** 152.

Cradick, T. J., Keck, K., Bradshaw, S., Jamieson, A. C. & McCaffrey, A. P. (2010) Zinc-finger Nucleases as a Novel Therapeutic Strategy for Targeting Hepatitis B Virus DNAs. *Mol Ther*.

Dailey, L., Ambrosetti, D., Mansukhani, A. & Basilico, C. (2005) Mechanisms underlying differential responses to FGF signaling. *Cytokine Growth Factor Rev,* 16**,** 233-47.

Darnell, J. E., Jr. (1997) STATs and gene regulation. *Science,* 277**,** 1630-5.

Data & Statistics, U. N. (2012) Breast Cancer: Incidence, mortality and survival. *http://www.ons.gov.uk/ons/rel/cancer-unit/breast-cancer-in-england/2010/sum-1.html,* 02.11.12.

de Jong, M. M., Nolte, I. M., te Meerman, G. J., van der Graaf, W. T., Oosterwijk, J. C., Kleibeuker, J. H., Schaapveld, M. & de Vries, E. G. (2002) Genes other than BRCA1 and BRCA2 involved in breast cancer susceptibility. *J Med Genet,* 39**,** 225-42.

Dellaire, G. & Chartrand, P. (1998) Direct evidence that transgene integration is random in murine cells, implying that naturally occurring double-strand breaks may be distributed similarly within the genome. *Radiat Res,* 149**,** 325-9.

Dent, R., Trudeau, M., Pritchard, K. I., Hanna, W. M., Kahn, H. K., Sawka, C. A., Lickley, L. A., Rawlinson, E., Sun, P. & Narod, S. A. (2007) Triple-negative breast cancer: clinical features and patterns of recurrence. *Clin Cancer Res,* 13**,** 4429-34.

Doyon, Y., Choi, V. M., Xia, D. F., Vo, T. D., Gregory, P. D. & Holmes, M. C. (2010) Transient cold shock enhances zinc-finger nuclease-mediated gene disruption. *Nat Methods,* 7**,** 459-60.

Doyon, Y., McCammon, J. M., Miller, J. C., Faraji, F., Ngo, C., Katibah, G. E., Amora, R., Hocking, T. D., Zhang, L., Rebar, E. J., Gregory, P. D., Urnov, F. D. & Amacher, S. L. (2008) Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases. *Nat Biotechnol,* 26**,** 702-8.

Doyon, Y., Vo, T. D., Mendel, M. C., Greenberg, S. G., Wang, J., Xia, D. F., Miller, J. C., Urnov, F. D., Gregory, P. D. & Holmes, M. C. (2011) Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. *Nat Methods,* 8**,** 74-9.

Easton, D. F., Pooley, K. A., Dunning, A. M., Pharoah, P. D., Thompson, D., Ballinger, D. G., Struewing, J. P., Morrison, J., Field, H., Luben, R., Wareham, N., Ahmed, S., Healey, C. S., Bowman, R., Meyer, K. B., Haiman, C. A., Kolonel, L. K., Henderson, B. E., Le Marchand, L., Brennan, P., Sangrajrang, S., Gaborieau, V., Odefrey, F., Shen, C. Y., Wu, P. E., Wang, H. C., Eccles, D., Evans, D. G., Peto, J., Fletcher, O., Johnson, N., Seal, S., Stratton, M. R., Rahman, N., Chenevix-Trench, G., Bojesen, S. E., Nordestgaard, B. G., Axelsson, C. K., Garcia-Closas, M., Brinton, L., Chanock, S., Lissowska, J., Peplonska, B., Nevanlinna, H., Fagerholm, R., Eerola, H., Kang, D., Yoo, K. Y., Noh, D. Y., Ahn, S. H., Hunter, D. J., Hankinson, S. E., Cox, D. G., Hall, P., Wedren, S., Liu, J., Low, Y. L., Bogdanova, N., Schurmann, P., Dork, T., Tollenaar, R. A., Jacobi, C. E., Devilee, P., Klijn, J. G., Sigurdson, A. J., Doody, M. M., Alexander, B. H., Zhang, J., Cox, A., Brock, I. W., MacPherson, G., Reed, M. W., Couch, F. J., Goode, E. L., Olson, J. E., Meijers-Heijboer, H., van den Ouweland, A., Uitterlinden, A., Rivadeneira, F., Milne, R. L., Ribas, G., Gonzalez-Neira, A., Benitez, J., Hopper, J. L., McCredie, M., Southey, M., Giles, G. G., Schroen, C., Justenhoven, C., Brauch, H., Hamann, U., Ko, Y. D., Spurdle, A. B., Beesley, J., Chen, X., Mannermaa, A., Kosma, V. M., Kataja, V., Hartikainen, J., Day, N. E*., et al.* (2007) Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature,* 447**,** 1087-93.

Eisen, A. & Irwin, E. (2002) Review: breast cancer is associated with a family history of the disease in first degree relatives. *Evid Based Nurs,* 5**,** 89.

Eissmann, M., Gutschner, T., Hammerle, M., Gunther, S., Caudron-Herger, M., Gross, M., Schirmacher, P., Rippe, K., Braun, T., Zornig, M. & Diederichs, S. (2012) Loss of the abundant nuclear non-coding RNA MALAT1 is compatible with life and development. *RNA Biol,* 9**,** 1076-87.

Enard, W., Khaitovich, P., Klose, J., Zollner, S., Heissig, F., Giavalisco, P., Nieselt-Struwe, K., Muchmore, E., Varki, A., Ravid, R., Doxiadis, G. M., Bontrop, R. E. & Paabo, S. (2002) Intra- and interspecific variation in primate gene expression patterns. *Science,* 296**,** 340-3.

Enattah, N. S., Sahi, T., Savilahti, E., Terwilliger, J. D., Peltonen, L. & Jarvela, I. (2002) Identification of a variant associated with adult-type hypolactasia. *Nat Genet,* 30**,** 233-7.

ENCODE (2012) ENCODE Consortium data base. *www.regulomedb.org,* 05.06.12**,** RegulomeDB is a database that annotates SNPs with known and predicted regulatory elements in the intergenic regions of the H. sapiens genome.

Center for Genomics and Personalized Medicine at Stanford University

Source of these data include public datasets from GEO, the ENCODE project, and published literature.

Ensembl (2010) FGFR2 variation, Ensembl Genome Browser website. *http://www.ensembl.org/Homo_sapiens/Transcript/Variation_Transcript/Table?db=core;g=ENSG00000066468;r=10:123237848-123357972;t=ENST00000358487#coding_sequence_variant_tablePanel,* 18.11.11.

Eswarakumar, V. P., Lax, I. & Schlessinger, J. (2005) Cellular signaling by fibroblast growth factor receptors. *Cytokine Growth Factor Rev,* 16**,** 139-49.

Fackenthal, J. D. & Olopade, O. I. (2007) Breast cancer risk associated with BRCA1 and BRCA2 in diverse populations. *Nat Rev Cancer,* 7**,** 937-48.

Farmer, H., McCabe, N., Lord, C. J., Tutt, A. N., Johnson, D. A., Richardson, T. B., Santarosa, M., Dillon, K. J., Hickson, I., Knights, C., Martin, N. M., Jackson, S. P., Smith, G. C. & Ashworth, A. (2005) Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature,* 434**,** 917-21.

Ferguson-Smith, A., Lin, S. P., Tsai, C. E., Youngson, N. & Tevendale, M. (2003) Genomic imprinting--insights from studies in mice. *Semin Cell Dev Biol,* 14**,** 43-9.

Fisher, B., Anderson, S., Bryant, J., Margolese, R. G., Deutsch, M., Fisher, E. R., Jeong, J. H. & Wolmark, N. (2002) Twenty-year follow-up of a randomized trial comparing total mastectomy, lumpectomy, and lumpectomy plus irradiation for the treatment of invasive breast cancer. *N Engl J Med,* 347**,** 1233-41.

Fisher, B., Redmond, C., Poisson, R., Margolese, R., Wolmark, N., Wickerham, L., Fisher, E., Deutsch, M., Caplan, R., Pilch, Y. & et al. (1989) Eight-year results of a randomized clinical trial comparing total mastectomy and lumpectomy with or without irradiation in the treatment of breast cancer. *N Engl J Med,* 320**,** 822-8.

Fong, P. C., Boss, D. S., Yap, T. A., Tutt, A., Wu, P., Mergui-Roelvink, M., Mortimer, P., Swaisland, H., Lau, A., O'Connor, M. J., Ashworth, A., Carmichael, J., Kaye, S. B., Schellens, J. H. & de Bono, J. S. (2009) Inhibition of poly(ADP-ribose) polymerase in tumors from BRCA mutation carriers. *N Engl J Med,* 361**,** 123-34.

Forrest P, C. J., Elton A, Evans K, Gravelle H (1986) Breast Cancer Screening. Report to the Health Ministers of England, Wales, Scotland and Northern Ireland

Francastel, C., Walters, M. C., Groudine, M. & Martin, D. I. (1999) A functional enhancer suppresses silencing of a transgene and prevents its localization close to centrometric heterochromatin. *Cell,* 99**,** 259-69.

Fu, J., Weise, A. M., Falany, J. L., Falany, C. N., Thibodeau, B. J., Miller, F. R., Kocarek, T. A. & Runge-Morris, M. (2010) Expression of estrogenicity genes in a lineage cell culture model of human breast cancer progression. *Breast Cancer Res Treat,* 120**,** 35-45.

Furdui, C. M., Lew, E. D., Schlessinger, J. & Anderson, K. S. (2006) Autophosphorylation of FGFR1 kinase is mediated by a sequential and precisely ordered reaction. *Mol Cell,* 21**,** 711-7.

Gabriel, R., Lombardo, A., Arens, A., Miller, J. C., Genovese, P., Kaeppel, C., Nowrouzi, A., Bartholomae, C. C., Wang, J., Friedman, G., Holmes, M. C., Gregory, P. D., Glimm, H., Schmidt, M., Naldini, L. & von Kalle, C. (2011) An unbiased genome-wide analysis of zinc-finger nuclease specificity. *Nat Biotechnol,* 29**,** 816-23.

Gartside, M. G., Chen, H., Ibrahimi, O. A., Byron, S. A., Curtis, A. V., Wellens, C. L., Bengston, A., Yudt, L. M., Eliseenkova, A. V., Ma, J., Curtin, J. A., Hyder, P., Harper, U. L., Riedesel, E., Mann, G. J., Trent, J. M., Bastian, B. C., Meltzer, P. S., Mohammadi, M. & Pollock, P. M. (2009) Loss-of-function fibroblast growth factor receptor-2 mutations in melanoma. *Mol Cancer Res,* 7**,** 41-54.

Genomes Project, C. (2010) A map of human genome variation from population-scale sequencing. *Nature,* 467**,** 1061-73.

Geurts, A. M., Cost, G. J., Freyvert, Y., Zeitler, B., Miller, J. C., Choi, V. M., Jenkins, S. S., Wood, A., Cui, X., Meng, X., Vincent, A., Lam, S., Michalkiewicz, M., Schilling, R., Foeckler, J., Kalloway, S., Weiler, H., Menoret, S., Anegon, I., Davis, G. D., Zhang, L., Rebar, E. J., Gregory, P. D., Urnov, F. D., Jacob, H. J. & Buelow, R. (2009) Knockout rats via embryo microinjection of zinc-finger nucleases. *Science,* 325**,** 433.

Gomez-Raposo, C., Zambrana Tevar, F., Sereno Moyano, M., Lopez Gomez, M. & Casado, E. (2010) Male breast cancer. *Cancer Treat Rev,* 36**,** 451-7.

Gotoh, N. (2008) Regulation of growth factor signaling by FRS2 family docking/scaffold adaptor proteins. *Cancer Sci,* 99**,** 1319-25.

Greenman, C., Stephens, P., Smith, R., Dalgliesh, G. L., Hunter, C., Bignell, G., Davies, H., Teague, J., Butler, A., Stevens, C., Edkins, S., O'Meara, S., Vastrik, I., Schmidt, E. E., Avis, T., Barthorpe, S., Bhamra, G., Buck, G., Choudhury, B., Clements, J., Cole, J., Dicks, E., Forbes, S., Gray, K., Halliday, K., Harrison, R., Hills, K., Hinton, J., Jenkinson, A., Jones, D., Menzies, A., Mironenko, T., Perry, J., Raine, K., Richardson, D., Shepherd, R., Small, A., Tofts, C., Varian, J., Webb, T., West, S., Widaa, S., Yates, A., Cahill, D. P., Louis, D. N., Goldstraw, P., Nicholson, A. G., Brasseur, F., Looijenga, L., Weber, B. L., Chiew, Y. E., DeFazio, A., Greaves, M. F., Green, A. R., Campbell, P., Birney, E., Easton, D. F., Chenevix-Trench, G., Tan, M. H., Khoo, S. K., Teh, B. T., Yuen, S. T., Leung, S. Y., Wooster, R., Futreal, P. A. & Stratton, M. R. (2007) Patterns of somatic mutation in human cancer genomes. *Nature,* 446**,** 153-8.

Grose, R. & Dickson, C. (2005) Fibroblast growth factor signaling in tumorigenesis. *Cytokine Growth Factor Rev,* 16**,** 179-86.

Gudas, J. M., Klein, R. C., Oka, M. & Cowan, K. H. (1995) Posttranscriptional regulation of the c-myb proto-oncogene in estrogen receptor-positive breast cancer cells. *Clin Cancer Res,* 1**,** 235-43.

Gurtu, V., Yan, G. & Zhang, G. (1996) IRES bicistronic expression vectors for efficient creation of stable mammalian cell lines. *Biochem Biophys Res Commun,* 229**,** 295-8.

Gutschner, T., Baas, M. & Diederichs, S. (2011) Noncoding RNA gene silencing through genomic integration of RNA destabilizing elements using zinc finger nucleases. *Genome Res,* 21**,** 1944-54.

Gutschner, T., Hammerle, M., Eissmann, M., Hsu, J., Kim, Y., Hung, G., Revenko, A. S., Arun, G., Stentrup, M., Gross, M., Zornig, M., Macleod, A. R., Spector, D. L. & Diederichs, S. (2012) The non-coding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells. *Cancer Res*.

Hall, J. M., Lee, M. K., Newman, B., Morrow, J. E., Anderson, L. A., Huey, B. & King, M. C. (1990) Linkage of early-onset familial breast cancer to chromosome 17q21. *Science,* 250**,** 1684-9.

Han, W., Woo, J. H., Yu, J. H., Lee, M. J., Moon, H. G., Kang, D. & Noh, D. Y. (2011) Common genetic variants associated with breast cancer in Korean women and differential susceptibility according to intrinsic subtype. *Cancer Epidemiol Biomarkers Prev,* 20**,** 793-8.

Hanahan, D. & Weinberg, R. A. (2000) The hallmarks of cancer. *Cell,* 100**,** 57-70.

Hanahan, D. & Weinberg, R. A. (2011) Hallmarks of cancer: the next generation. *Cell,* 144**,** 646-74.

Heel, R. C., Brogden, R. N., Speight, T. M. & Avery, G. S. (1978) Tamoxifen: a review of its pharmacological properties and therapeutic use in the treatment of breast cancer. *Drugs,* 16**,** 1-24.

Heiskanen, M., Kononen, J., Barlund, M., Torhorst, J., Sauter, G., Kallioniemi, A. & Kallioniemi, O. (2001) CGH, cDNA and tissue microarray analyses implicate FGFR2 amplification in a small subset of breast tumors. *Anal Cell Pathol,* 22**,** 229-34.

Hindorff, L. A., Sethupathy, P., Junkins, H. A., Ramos, E. M., Mehta, J. P., Collins, F. S. & Manolio, T. A. (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A,* 106**,** 9362-7.

Hishikawa, Y., Tamaru, N., Ejima, K., Hayashi, T. & Koji, T. (2004) Expression of keratinocyte growth factor and its receptor in human breast cancer: its inhibitory role in the induction of apoptosis possibly through the overexpression of Bcl-2. *Arch Histol Cytol,* 67**,** 455-64.

Hockemeyer, D., Soldner, F., Beard, C., Gao, Q., Mitalipova, M., DeKelver, R. C., Katibah, G. E., Amora, R., Boydston, E. A., Zeitler, B., Meng, X., Miller, J. C., Zhang, L., Rebar, E. J., Gregory, P. D., Urnov, F. D. & Jaenisch, R. (2009) Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. *Nat Biotechnol,* 27**,** 851-7.

Holliday, D. L., Brouilette, K. T., Markert, A., Gordon, L. A. & Jones, J. L. (2009) Novel multicellular organotypic models of normal and malignant breast: tools for dissecting the role of the microenvironment in breast cancer progression. *Breast Cancer Res,* 11**,** R3.

Holt, N., Wang, J., Kim, K., Friedman, G., Wang, X., Taupin, V., Crooks, G. M., Kohn, D. B., Gregory, P. D., Holmes, M. C. & Cannon, P. M. (2010) Human hematopoietic

stem/progenitor cells modified by zinc-finger nucleases targeted to CCR5 control HIV-1 in vivo. *Nat Biotechnol,* 28**,** 839-47.

Hopper, J. L. & Carlin, J. B. (1992) Familial aggregation of a disease consequent upon correlation between relatives in a risk factor measured on a continuous scale. *Am J Epidemiol,* 136**,** 1138-47.

Hughes, A. L., Packer, B., Welch, R., Chanock, S. J. & Yeager, M. (2005) High level of functional polymorphism indicates a unique role of natural selection at human immune system loci. *Immunogenetics,* 57**,** 821-7.

Huijts, P. E., van Dongen, M., de Goeij, M. C., van Moolenbroek, A. J., Blanken, F., Vreeswijk, M. P., de Kruijf, E. M., Mesker, W. E., van Zwet, E. W., Tollenaar, R. A., Smit, V. T., van Asperen, C. J. & Devilee, P. (2011) Allele-specific regulation of FGFR2 expression is cell type-dependent and may increase breast cancer risk through a paracrine stimulus involving FGF10. *Breast Cancer Res,* 13**,** R72.

Hunter, D. J., Kraft, P., Jacobs, K. B., Cox, D. G., Yeager, M., Hankinson, S. E., Wacholder, S., Wang, Z., Welch, R., Hutchinson, A., Wang, J., Yu, K., Chatterjee, N., Orr, N., Willett, W. C., Colditz, G. A., Ziegler, R. G., Berg, C. D., Buys, S. S., McCarty, C. A., Feigelson, H. S., Calle, E. E., Thun, M. J., Hayes, R. B., Tucker, M., Gerhard, D. S., Fraumeni, J. F., Jr., Hoover, R. N., Thomas, G. & Chanock, S. J. (2007) A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat Genet,* 39**,** 870-4.

Hurtado, A., Holmes, K. A., Ross-Innes, C. S., Schmidt, D. & Carroll, J. S. (2010) FOXA1 is a key determinant of estrogen receptor function and endocrine response. *Nat Genet,* 43**,** 27-33.

Hutter, C. M., Young, A. M., Ochs-Balcom, H. M., Carty, C. L., Wang, T., Chen, C. T., Rohan, T. E., Kooperberg, C. & Peters, U. (2011) Replication of breast cancer GWAS susceptibility loci in the Women's Health Initiative African American SHARe Study. *Cancer Epidemiol Biomarkers Prev,* 20**,** 1950-9.

Hynes, N. E. & Watson, C. J. (2010) Mammary gland growth factors: roles in normal development and in cancer. *Cold Spring Harb Perspect Biol,* 2**,** a003186.

Inman, C. K., Li, N. & Shore, P. (2005) Oct-1 counteracts autoinhibition of Runx2 DNA binding to form a novel Runx2/Oct-1 complex on the promoter of the mammary gland-specific gene beta-casein. *Mol Cell Biol,* 25**,** 3182-93.

Itoh, N. & Ornitz, D. M. (2008) Functional evolutionary history of the mouse Fgf gene family. *Dev Dyn,* 237**,** 18-27.

James, J. J., Evans, A. J., Pinder, S. E., Gutteridge, E., Cheung, K. L., Chan, S. & Robertson, J. F. (2003) Bone metastases from breast carcinoma: histopathological - radiological correlations and prognostic features. *Br J Cancer,* 89**,** 660-5.

Jenne, D. E., Reimann, H., Nezu, J., Friedel, W., Loff, S., Jeschke, R., Muller, O., Back, W. & Zimmer, M. (1998) Peutz-Jeghers syndrome is caused by mutations in a novel serine threonine kinase. *Nat Genet,* 18**,** 38-43.

Jin, G., Sun, J., Isaacs, S. D., Wiley, K. E., Kim, S. T., Chu, L. W., Zhang, Z., Zhao, H., Zheng, S. L., Isaacs, W. B. & Xu, J. (2011) Human polymorphisms at long non-coding RNAs (lncRNAs) and association with prostate cancer risk. *Carcinogenesis,* 32**,** 1655-9.

Kadota, M., Sato, M., Duncan, B., Ooshima, A., Yang, H. H., Diaz-Meyer, N., Gere, S., Kageyama, S., Fukuoka, J., Nagata, T., Tsukada, K., Dunn, B. K., Wakefield, L. M. & Lee, M. P. (2009) Identification of novel gene amplifications in breast cancer and coexistence of gene amplification with an activating mutation of PIK3CA. *Cancer Res,* 69**,** 7357-65.

Kadota, M., Yang, H. H., Gomez, B., Sato, M., Clifford, R. J., Meerzaman, D., Dunn, B. K., Wakefield, L. M. & Lee, M. P. (2010) Delineating genetic alterations for tumor progression in the MCF10A series of breast cancer cell lines. *PLoS One,* 5**,** e9201.

Kelsey, J. L. & Berkowitz, G. S. (1988) Breast cancer epidemiology. *Cancer Res,* 48**,** 5615-23.

Keydar, I., Chen, L., Karby, S., Weiss, F. R., Delarea, J., Radu, M., Chaitcik, S. & Brenner, H. J. (1979) Establishment and characterization of a cell line of human breast carcinoma origin. *Eur J Cancer,* 15**,** 659-70.

Kim, H., Um, E., Cho, S. R., Jung, C., Kim, H. & Kim, J. S. (2011) Surrogate reporters for enrichment of cells with nuclease-induced mutations. *Nat Methods,* 8**,** 941-3.

Kim, J., Petz, L. N., Ziegler, Y. S., Wood, J. R., Potthoff, S. J. & Nardulli, A. M. (2000) Regulation of the estrogen-responsive pS2 gene in MCF-7 human breast cancer cells. *J Steroid Biochem Mol Biol,* 74**,** 157-68.

Kim, Y. G., Cha, J. & Chandrasegaran, S. (1996) Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc Natl Acad Sci U S A,* 93**,** 1156-60.

King, M. C. & Wilson, A. C. (1975) Evolution at two levels in humans and chimpanzees. *Science,* 188**,** 107-16.

Kouhara, H., Hadari, Y. R., Spivak-Kroizman, T., Schilling, J., Bar-Sagi, D., Lax, I. & Schlessinger, J. (1997) A lipid-anchored Grb2-binding protein that links FGF-receptor activation to the Ras/MAPK signaling pathway. *Cell,* 89**,** 693-702.

Koziczak, M., Holbro, T. & Hynes, N. E. (2004) Blocking of FGFR signaling inhibits breast cancer cell proliferation through downregulation of D-type cyclins. *Oncogene,* 23**,** 3501-8.

Krag, D. N., Anderson, S. J., Julian, T. B., Brown, A. M., Harlow, S. P., Costantino, J. P., Ashikaga, T., Weaver, D. L., Mamounas, E. P., Jalovec, L. M., Frazier, T. G., Noyes, R. D., Robidoux, A., Scarth, H. M. & Wolmark, N. (2010) Sentinel-lymph-node resection compared with conventional axillary-lymph-node dissection in clinically node-negative patients with breast cancer: overall survival findings from the NSABP B-32 randomised phase 3 trial. *Lancet Oncol,* 11**,** 927-33.

Kufe, D., Pollock, R., Weichselbaum, R., Bast, R., Gansler, T., Holland, J. & Frei, E. (2003) *Holland-Frei Cancer Medicine, 6th edition*, Hamilton (ON): BC Decker.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M.,

Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., McPherson, J. D., Marra, M. A., Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin, K. H., Gish, W. R., Chissoe, S. L., Wendl, M. C., Delehaunty, K. D., Miner, T. L., Delehaunty, A., Kramer, J. B., Cook, L. L., Fulton, R. S., Johnson, D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J. F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M*., et al.* (2001) Initial sequencing and analysis of the human genome. *Nature,* 409**,** 860-921.

Ledford, H. (2011) Targeted gene editing enters clinic. *Nature,* 471**,** 16.

Lee, H. J., Kim, E. & Kim, J. S. (2009) Targeted chromosomal deletions in human cells using zinc finger nucleases. *Genome Res,* 20**,** 81-9.

Lee, P. H. & Shatkay, H. (2008) F-SNP: computationally predicted functional SNPs for disease association studies. *Nucleic Acids Res,* 36**,** D820-4.

Li, W. H. & Sadler, L. A. (1991) Low nucleotide diversity in man. *Genetics,* 129**,** 513-23.

Lieber, M. R. (2008) The mechanism of human nonhomologous DNA end joining. *J Biol Chem,* 283**,** 1-5.

Lieber, M. R. (2010) The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu Rev Biochem,* 79**,** 181-211.

Lin, C. C., Melo, F. A., Ghosh, R., Suen, K. M., Stagg, L. J., Kirkpatrick, J., Arold, S. T., Ahmed, Z. & Ladbury, J. E. (2012) Inhibition of basal FGF receptor signaling by dimeric Grb2. *Cell,* 149**,** 1514-24.

Liu, P. Q., Chan, E. M., Cost, G. J., Zhang, L., Wang, J., Miller, J. C., Guschin, D. Y., Reik, A., Holmes, M. C., Mott, J. E., Collingwood, T. N. & Gregory, P. D. (2010) Generation of a triple-gene knockout mammalian cell line using engineered zinc-finger nucleases. *Biotechnol Bioeng,* 106**,** 97-105.

Liu, R., Paxton, W. A., Choe, S., Ceradini, D., Martin, S. R., Horuk, R., MacDonald, M. E., Stuhlmann, H., Koup, R. A. & Landau, N. R. (1996) Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. *Cell,* 86**,** 367-77.

Livak, K. J. & Schmittgen, T. D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods,* 25**,** 402-8.

Lloyd, A., Plaisier, C. L., Carroll, D. & Drews, G. N. (2005) Targeted mutagenesis using zinc-finger nucleases in Arabidopsis. *Proc Natl Acad Sci U S A,* 102**,** 2232-7.

Long, J., Shu, X. O., Cai, Q., Gao, Y. T., Zheng, Y., Li, G., Li, C., Gu, K., Wen, W., Xiang, Y. B., Lu, W. & Zheng, W. (2010) Evaluation of breast cancer susceptibility loci in Chinese women. *Cancer Epidemiol Biomarkers Prev,* 19**,** 2357-65.

Lower, E. E., Glass, E. L., Bradley, D. A., Blau, R. & Heffelfinger, S. (2005) Impact of metastatic estrogen receptor and progesterone receptor status on survival. *Breast Cancer Res Treat,* 90**,** 65-70.

Lu, P., Ewald, A. J., Martin, G. R. & Werb, Z. (2008) Genetic mosaic analysis reveals FGF receptor 2 function in terminal end buds during mammary gland branching morphogenesis. *Dev Biol,* 321**,** 77-87.

Mailleux, A. A., Spencer-Dene, B., Dillon, C., Ndiaye, D., Savona-Baron, C., Itoh, N., Kato, S., Dickson, C., Thiery, J. P. & Bellusci, S. (2002) Role of FGF10/FGFR2b signaling during mammary gland development in the mouse embryo. *Development,* 129**,** 53-60.

Mani, M., Kandavelou, K., Dy, F. J., Durai, S. & Chandrasegaran, S. (2005) Design, engineering, and characterization of zinc finger nucleases. *Biochem Biophys Res Commun,* 335**,** 447-57.

Marsh, D. J., Kum, J. B., Lunetta, K. L., Bennett, M. J., Gorlin, R. J., Ahmed, S. F., Bodurtha, J., Crowe, C., Curtis, M. A., Dasouki, M., Dunn, T., Feit, H., Geraghty, M. T., Graham, J. M., Jr., Hodgson, S. V., Hunter, A., Korf, B. R., Manchester, D., Miesfeldt, S., Murday, V. A., Nathanson, K. L., Parisi, M., Pober, B., Romano, C., Eng, C. & et al. (1999) PTEN mutation spectrum and genotype-phenotype correlations in Bannayan-Riley-Ruvalcaba syndrome suggest a single entity with Cowden syndrome. *Hum Mol Genet,* 8**,** 1461-72.

Mavaddat, N., Antoniou, A. C., Easton, D. F. & Garcia-Closas, M. (2010) Genetic susceptibility to breast cancer. *Mol Oncol,* 4**,** 174-91.

Meijers-Heijboer, H., van den Ouweland, A., Klijn, J., Wasielewski, M., de Snoo, A., Oldenburg, R., Hollestelle, A., Houben, M., Crepin, E., van Veghel-Plandsoen, M., Elstrodt, F., van Duijn, C., Bartels, C., Meijers, C., Schutte, M., McGuffog, L., Thompson, D., Easton, D., Sodha, N., Seal, S., Barfoot, R., Mangion, J., Chang-Claude, J., Eccles, D., Eeles, R., Evans, D. G., Houlston, R., Murday, V., Narod, S., Peretz, T., Peto, J., Phelan, C., Zhang, H. X., Szabo, C., Devilee, P., Goldgar, D., Futreal, P. A., Nathanson, K. L., Weber, B., Rahman, N., Stratton, M. R. & Consortium, C. H.-B. C. (2002) Low-penetrance susceptibility to breast cancer due to CHEK2(*)1100delC in noncarriers of BRCA1 or BRCA2 mutations. *Nat Genet,* 31**,** 55-9.

Meyer, K. B., Maia, A. T., O'Reilly, M., Teschendorff, A. E., Chin, S. F., Caldas, C. & Ponder, B. A. (2008) Allele-specific up-regulation of FGFR2 increases susceptibility to breast cancer. *PLoS Biol,* 6**,** e108.

Michailidou K, H. P., Gonzalez-Neira A, Ghoussaini M et al (2013) Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nature Genetics*.

Mignatti, P., Morimoto, T. & Rifkin, D. B. (1992) Basic fibroblast growth factor, a protein devoid of secretory signal sequence, is released by cells via a pathway independent of the endoplasmic reticulum-Golgi complex. *J Cell Physiol,* 151**,** 81-93.

Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P. A., Harshman, K., Tavtigian, S., Liu, Q., Cochran, C., Bennett, L. M., Ding, W. & et al. (1994) A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science,* 266**,** 66-71.

Milani, L., Gupta, M., Andersen, M., Dhar, S., Fryknas, M., Isaksson, A., Larsson, R. & Syvanen, A. C. (2007) Allelic imbalance in gene expression as a guide to cis-acting regulatory single nucleotide polymorphisms in cancer cells. *Nucleic Acids Res,* 35**,** e34.

Miller, F. R., Santner, S. J., Tait, L. & Dawson, P. J. (2000) MCF10DCIS.com xenograft model of human comedo ductal carcinoma in situ. *J Natl Cancer Inst,* 92**,** 1185-6.

Miller, J. C., Holmes, M. C., Wang, J., Guschin, D. Y., Lee, Y. L., Rupniewski, I., Beausejour, C. M., Waite, A. J., Wang, N. S., Kim, K. A., Gregory, P. D., Pabo, C. O. & Rebar, E. J. (2007) An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat Biotechnol,* 25**,** 778-85.

Moehle, E. A., Rock, J. M., Lee, Y. L., Jouvenot, Y., DeKelver, R. C., Gregory, P. D., Urnov, F. D. & Holmes, M. C. (2007) Targeted gene addition into a specified location in the human genome using designed zinc finger nucleases. *Proc Natl Acad Sci U S A,* 104**,** 3055-60.

Mohammadi, M., Dikic, I., Sorokin, A., Burgess, W. H., Jaye, M. & Schlessinger, J. (1996) Identification of six novel autophosphorylation sites on fibroblast growth factor receptor 1 and elucidation of their importance in receptor activation and signal transduction. *Mol Cell Biol,* 16**,** 977-89.

Mohammadi, M., Honegger, A. M., Rotin, D., Fischer, R., Bellot, F., Li, W., Dionne, C. A., Jaye, M., Rubinstein, M. & Schlessinger, J. (1991) A tyrosine-phosphorylated carboxy-terminal peptide of the fibroblast growth factor receptor (Flg) is a binding site for the SH2 domain of phospholipase C-gamma 1. *Mol Cell Biol,* 11**,** 5068-78.

Morley, M., Molony, C. M., Weber, T. M., Devlin, J. L., Ewens, K. G., Spielman, R. S. & Cheung, V. G. (2004) Genetic analysis of genome-wide variation in human gene expression. *Nature,* 430**,** 743-7.

Motulsky, A. G. (2006) Genetics of complex diseases. *J Zhejiang Univ Sci B,* 7**,** 167-8.

Moynahan, M. E. & Jasin, M. (2010) Mitotic homologous recombination maintains genomic stability and suppresses tumorigenesis. *Nat Rev Mol Cell Biol,* 11**,** 196-207.

Muller, H., Bracken, A. P., Vernell, R., Moroni, M. C., Christians, F., Grassilli, E., Prosperini, E., Vigo, E., Oliner, J. D. & Helin, K. (2001) E2Fs regulate the expression of genes involved in differentiation, development, proliferation, and apoptosis. *Genes Dev,* 15**,** 267-85.

Mulligan, A. M., Couch, F. J., Barrowdale, D., Domchek, S. M., Eccles, D., Nevanlinna, H., Ramus, S. J., Robson, M., Sherman, M., Spurdle, A. B., Wappenschmidt, B., Lee, A., McGuffog, L., Healey, S., Sinilnikova, O. M., Janavicius, R., Hansen, T., Nielsen, F. C., Ejlertsen, B., Osorio, A., Munoz-Repeto, I., Duran, M., Godino, J., Pertesi, M., Benitez, J., Peterlongo, P., Manoukian, S., Peissel, B., Zaffaroni, D., Cattaneo, E., Bonanni, B., Viel, A., Pasini, B., Papi, L., Ottini, L., Savarese, A., Bernard, L., Radice, P., Hamann, U., Verheus, M., Meijers-Heijboer, H. E., Wijnen, J., Gomez Garcia, E. B., Nelen, M. R., Kets, C. M., Seynaeve, C., Tilanus-Linthorst, M. M., van der Luijt, R. B., van Os, T., Rookus, M., Frost, D., Jones, J. L., Evans, D. G., Lalloo, F., Eeles, R., Izatt, L., Adlard, J., Davidson, R., Cook, J., Donaldson, A., Dorkins, H., Gregory, H., Eason, J., Houghton, C., Barwell, J., Side, L. E., McCann, E., Murray, A., Peock, S.,

Godwin, A. K., Schmutzler, R. K., Rhiem, K., Engel, C., Meindl, A., Ruehl, I., Arnold, N., Niederacher, D., Sutter, C., Deissler, H., Gadzicki, D., Kast, K., Preisler-Adams, S., Varon-Mateeva, R., Schoenbuchner, I., Fiebig, B., Heinritz, W., Schafer, D., Gevensleben, H., Caux-Moncoutier, V., Fassy-Colcombet, M., Cornelis, F., Mazoyer, S., Leone, M., Boutry-Kryza, N., Hardouin, A., Berthet, P., Muller, D., Fricker, J. P., Mortemousque, I., Pujol, P*., et al.* (2011) Common breast cancer susceptibility alleles are associated with tumour subtypes in BRCA1 and BRCA2 mutation carriers: results from the Consortium of Investigators of Modifiers of BRCA1/2. *Breast Cancer Res,* 13**,** R110.

Neve, R. M., Chin, K., Fridlyand, J., Yeh, J., Baehner, F. L., Fevr, T., Clark, L., Bayani, N., Coppe, J. P., Tong, F., Speed, T., Spellman, P. T., DeVries, S., Lapuk, A., Wang, N. J., Kuo, W. L., Stilwell, J. L., Pinkel, D., Albertson, D. G., Waldman, F. M., McCormick, F., Dickson, R. B., Johnson, M. D., Lippman, M., Ethier, S., Gazdar, A. & Gray, J. W. (2006) A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell,* 10**,** 515-27.

Ng, K. P., Hillmer, A. M., Chuah, C. T., Juan, W. C., Ko, T. K., Teo, A. S., Ariyaratne, P. N., Takahashi, N., Sawada, K., Fei, Y., Soh, S., Lee, W. H., Huang, J. W., Allen, J. C., Jr., Woo, X. Y., Nagarajan, N., Kumar, V., Thalamuthu, A., Poh, W. T., Ang, A. L., Mya, H. T., How, G. F., Yang, L. Y., Koh, L. P., Chowbay, B., Chang, C. T., Nadarajan, V. S., Chng, W. J., Than, H., Lim, L. C., Goh, Y. T., Zhang, S., Poh, D., Tan, P., Seet, J. E., Ang, M. K., Chau, N. M., Ng, Q. S., Tan, D. S., Soda, M., Isobe, K., Nothen, M. M., Wong, T. Y., Shahab, A., Ruan, X., Cacheux-Rataboul, V., Sung, W. K., Tan, E. H., Yatabe, Y., Mano, H., Soo, R. A., Chin, T. M., Lim, W. T., Ruan, Y. & Ong, S. T. (2012) A common BIM deletion polymorphism mediates intrinsic resistance and inferior responses to tyrosine kinase inhibitors in cancer. *Nat Med,* 18**,** 521-8.

Nickerson, D. A., Taylor, S. L., Weiss, K. M., Clark, A. G., Hutchinson, R. G., Stengard, J., Salomaa, V., Vartiainen, E., Boerwinkle, E. & Sing, C. F. (1998) DNA sequence diversity in a 9.7-kb region of the human lipoprotein lipase gene. *Nat Genet,* 19**,** 233-40.

Nystrom, M. L., Thomas, G. J., Stone, M., Mackenzie, I. C., Hart, I. R. & Marshall, J. F. (2005) Development of a quantitative method to analyse tumour cell invasion in organotypic culture. *J Pathol,* 205**,** 468-75.

Ohlsson, R., Tycko, B. & Sapienza, C. (1998) Monoallelic expression: 'there can only be one'. *Trends Genet,* 14**,** 435-8.

Olds, L. C. & Sibley, E. (2003) Lactase persistence DNA variant enhances lactase promoter activity in vitro: functional role as a cis regulatory element. *Hum Mol Genet,* 12**,** 2333-40.

Ong, S. H., Hadari, Y. R., Gotoh, N., Guy, G. R., Schlessinger, J. & Lax, I. (2001) Stimulation of phosphatidylinositol 3-kinase by fibroblast growth factor receptors is mediated by coordinated recruitment of multiple docking proteins. *Proc Natl Acad Sci U S A,* 98**,** 6074-9.

Ornitz, D. M. & Itoh, N. (2001) Fibroblast growth factors. *Genome Biol,* 2**,** REVIEWS3005.

Ornitz, D. M., Xu, J., Colvin, J. S., McEwen, D. G., MacArthur, C. A., Coulier, F., Gao, G. & Goldfarb, M. (1996) Receptor specificity of the fibroblast growth factor family. *J Biol Chem,* 271**,** 15292-7.

Orr-Urtreger, A., Bedford, M. T., Burakova, T., Arman, E., Zimmer, Y., Yayon, A., Givol, D. & Lonai, P. (1993) Developmental localization of the splicing alternatives of fibroblast growth factor receptor-2 (FGFR2). *Dev Biol,* 158**,** 475-86.

Orr, N., Cooke, R., Jones, M., Fletcher, O., Dudbridge, F., Chilcott-Burns, S., Tomczyk, K., Broderick, P., Houlston, R., Ashworth, A. & Swerdlow, A. (2011) Genetic variants at chromosomes 2q35, 5p12, 6q25.1, 10q26.13, and 16q12.1 influence the risk of breast cancer in men. *PLoS Genet,* 7**,** e1002290.

Orr, N., Lemnrau, A., Cooke, R., Fletcher, O., Tomczyk, K., Jones, M., Johnson, N., Lord, C. J., Mitsopoulos, C., Zvelebil, M., McDade, S. S., Buck, G., Blancher, C., Consortium, K. C., Trainer, A. H., James, P. A., Bojesen, S. E., Bokmand, S., Nevanlinna, H., Mattson, J., Friedman, E., Laitman, Y., Palli, D., Masala, G., Zanna, I., Ottini, L., Giannini, G., Hollestelle, A., Ouweland, A. M., Novakovic, S., Krajc, M., Gago-Dominguez, M., Castelao, J. E., Olsson, H., Hedenfalk, I., Easton, D. F., Pharoah, P. D., Dunning, A. M., Bishop, D. T., Neuhausen, S. L., Steele, L., Houlston, R. S., Garcia-Closas, M., Ashworth, A. & Swerdlow, A. J. (2012) Genome-wide association study identifies a common variant in RAD51B associated with male breast cancer risk. *Nat Genet,* 44**,** 1182-4.

Park, J. G., Oie, H. K., Sugarbaker, P. H., Henslee, J. G., Chen, T. R., Johnson, B. E. & Gazdar, A. (1987) Characteristics of cell lines established from human colorectal carcinoma. *Cancer Res,* 47**,** 6710-8.

Park, Y. Y., Park, E. S., Kim, S. B., Kim, S. C., Sohn, B. H., Chu, I. S., Jeong, W., Mills, G. B., Byers, L. A. & Lee, J. S. (2012) Development and validation of a prognostic gene-expression signature for lung adenocarcinoma. *PLoS One,* 7**,** e44225.

Parkin, D. M. (2011) 1. The fraction of cancer attributable to lifestyle and environmental factors in the UK in 2010. *Br J Cancer,* 105 Suppl 2**,** S2-5.

Parsa, S., Ramasamy, S. K., De Langhe, S., Gupte, V. V., Haigh, J. J., Medina, D. & Bellusci, S. (2008) Terminal end bud maintenance in mammary gland is dependent upon FGFR2b signaling. *Dev Biol,* 317**,** 121-31.

Pastinen, T. & Hudson, T. J. (2004) Cis-acting regulatory variation in the human genome. *Science,* 306**,** 647-50.

Perez, E. E., Wang, J., Miller, J. C., Jouvenot, Y., Kim, K. A., Liu, O., Wang, N., Lee, G., Bartsevich, V. V., Lee, Y. L., Guschin, D. Y., Rupniewski, I., Waite, A. J., Carpenito, C., Carroll, R. G., Orange, J. S., Urnov, F. D., Rebar, E. J., Ando, D., Gregory, P. D., Riley, J. L., Holmes, M. C. & June, C. H. (2008) Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. *Nat Biotechnol,* 26**,** 808-16.

Perou, C. M., Sorlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslen, L. A., Fluge, O., Pergamenschikov, A., Williams, C., Zhu, S. X., Lonning, P. E., Borresen-Dale, A. L., Brown, P. O. & Botstein, D. (2000) Molecular portraits of human breast tumours. *Nature,* 406**,** 747-52.

Peto, J., Collins, N., Barfoot, R., Seal, S., Warren, W., Rahman, N., Easton, D. F., Evans, C., Deacon, J. & Stratton, M. R. (1999) Prevalence of BRCA1 and BRCA2 gene mutations in patients with early-onset breast cancer. *J Natl Cancer Inst,* 91**,** 943-9.

Plotnikov, A. N., Hubbard, S. R., Schlessinger, J. & Mohammadi, M. (2000) Crystal structures of two FGF-FGFR complexes reveal the determinants of ligand-receptor specificity. *Cell,* 101**,** 413-24.

Pollock, P. M., Gartside, M. G., Dejeza, L. C., Powell, M. A., Mallon, M. A., Davies, H., Mohammadi, M., Futreal, P. A., Stratton, M. R., Trent, J. M. & Goodfellow, P. J. (2007) Frequent activating FGFR2 mutations in endometrial carcinomas parallel germline mutations associated with craniosynostosis and skeletal dysplasia syndromes. *Oncogene,* 26**,** 7158-62.

Porteus, M. H. & Baltimore, D. (2003) Chimeric nucleases stimulate gene targeting in human cells. *Science,* 300**,** 763.

Poulter, M., Hollox, E., Harvey, C. B., Mulcare, C., Peuhkuri, K., Kajander, K., Sarner, M., Korpela, R. & Swallow, D. M. (2003) The causal element for the lactase persistence/non-persistence polymorphism is located in a 1 Mb region of linkage disequilibrium in Europeans. *Ann Hum Genet,* 67**,** 298-311.

Raffioni, S., Thomas, D., Foehr, E. D., Thompson, L. M. & Bradshaw, R. A. (1999) Comparison of the intracellular signaling responses by three chimeric fibroblast growth factor receptors in PC12 cells. *Proc Natl Acad Sci U S A,* 96**,** 7178-83.

Ramakrishna, S., Kim, Y. H. & Kim, H. (2013) Stability of Zinc Finger Nuclease Protein Is Enhanced by the Proteasome Inhibitor MG132. *PLoS One,* 8**,** e54282.

Rapraeger, A. C., Krufka, A. & Olwin, B. B. (1991) Requirement of heparan sulfate for bFGF-mediated fibroblast growth and myoblast differentiation. *Science,* 252**,** 1705-8.

Raskin, L., Pinchev, M., Arad, C., Lejbkowicz, F., Tamir, A., Rennert, H. S., Rennert, G. & Gruber, S. B. (2008) FGFR2 is a breast cancer susceptibility gene in Jewish and Arab Israeli populations. *Cancer Epidemiol Biomarkers Prev,* 17**,** 1060-5.

Razin, S. V., Borunova, V. V., Maksimenko, O. G. & Kantidze, O. L. (2012) Cys2His2 zinc finger protein family: classification, functions, and major members. *Biochemistry (Mosc),* 77**,** 217-26.

Recillas-Targa, F. (2006) Multiple strategies for gene transfer, expression, knockdown, and chromatin influence in mammalian cell lines and transgenic animals. *Mol Biotechnol,* 34**,** 337-54.

Ritz, J., Martin, J. S. & Laederach, A. (2012) Evaluating our ability to predict the structural disruption of RNA by SNPs. *BMC Genomics,* 13 Suppl 4**,** S6.

Robinson, G. W. (2007) Cooperation of signalling pathways in embryonic mammary gland development. *Nat Rev Genet,* 8**,** 963-72.

Romond, E. H., Perez, E. A., Bryant, J., Suman, V. J., Geyer, C. E., Jr., Davidson, N. E., Tan-Chiu, E., Martino, S., Paik, S., Kaufman, P. A., Swain, S. M., Pisansky, T. M., Fehrenbacher, L., Kutteh, L. A., Vogel, V. G., Visscher, D. W., Yothers, G., Jenkins, R. B., Brown, A. M., Dakhil, S. R., Mamounas, E. P., Lingle, W. L., Klein, P. M., Ingle, J. N. & Wolmark, N. (2005) Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer. *N Engl J Med,* 353**,** 1673-84.

Ropiquet, F., Huguenin, S., Villette, J. M., Ronfle, V., Le Brun, G., Maitland, N. J., Cussenot, O., Fiet, J. & Berthon, P. (1999) FGF7/KGF triggers cell transformation and invasion on immortalised human prostatic epithelial PNT1A cells. *Int J Cancer,* 82**,** 237-43.

Ross-Innes, C. S., Stark, R., Teschendorff, A. E., Holmes, K. A., Ali, H. R., Dunning, M. J., Brown, G. D., Gojis, O., Ellis, I. O., Green, A. R., Ali, S., Chin, S. F., Palmieri, C., Caldas, C. & Carroll, J. S. (2012) Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature,* 481**,** 389-93.

Sachidanandam, R., Weissman, D., Schmidt, S. C., Kakol, J. M., Stein, L. D., Marth, G., Sherry, S., Mullikin, J. C., Mortimore, B. J., Willey, D. L., Hunt, S. E., Cole, C. G., Coggill, P. C., Rice, C. M., Ning, Z., Rogers, J., Bentley, D. R., Kwok, P. Y., Mardis, E. R., Yeh, R. T., Schultz, B., Cook, L., Davenport, R., Dante, M., Fulton, L., Hillier, L., Waterston, R. H., McPherson, J. D., Gilman, B., Schaffner, S., Van Etten, W. J., Reich, D., Higgins, J., Daly, M. J., Blumenstiel, B., Baldwin, J., Stange-Thomann, N., Zody, M. C., Linton, L., Lander, E. S. & Altshuler, D. (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature,* 409**,** 928-33.

Santner, S. J., Dawson, P. J., Tait, L., Soule, H. D., Eliason, J., Mohamed, A. N., Wolman, S. R., Heppner, G. H. & Miller, F. R. (2001) Malignant MCF10CA1 cell lines derived from premalignant human breast epithelial MCF10AT cells. *Breast Cancer Res Treat,* 65**,** 101-10.

Sanyal, A., Lajoie, B. R., Jain, G. & Dekker, J. (2012) The long-range interaction landscape of gene promoters. *Nature,* 489**,** 109-13.

Schadt, E. E., Monks, S. A., Drake, T. A., Lusis, A. J., Che, N., Colinayo, V., Ruff, T. G., Milligan, S. B., Lamb, J. R., Cavet, G., Linsley, P. S., Mao, M., Stoughton, R. B. & Friend, S. H. (2003) Genetics of gene expression surveyed in maize, mouse and man. *Nature,* 422**,** 297-302.

Schwertfeger, K. L. (2009) Fibroblast growth factors in development and cancer: insights from the mammary and prostate glands. *Curr Drug Targets,* 10**,** 632-44.

Scribner, K. C., Behbod, F. & Porter, W. W. (2012) Regulation of DCIS to invasive breast cancer progression by Singleminded-2s (SIM2s). *Oncogene*.

Sedivy, J. M. & Sharp, P. A. (1989) Positive genetic selection for gene disruption in mammalian cells by homologous recombination. *Proc Natl Acad Sci U S A,* 86**,** 227-31.

Serra, S., Zheng, L., Hassan, M., Pham, A. T., Woodhouse, L. J., Yao, J. C., Ezzat, S. & Asa, S. L. (2012) The FGFR4-G388R Single Nucleotide Polymorphism Alters Pancreatic Neuroendocrine Tumor Progression and Response to mTOR Inhibition Therapy. *Cancer Res*.

Shan, J., Mahfoudh, W., Dsouza, S. P., Hassen, E., Bouaouina, N., Abdelhak, S., Benhadjayed, A., Memmi, H., Mathew, R. A., Aigha, II, Gabbouj, S., Remadi, Y. & Chouchane, L. (2012) Genome-Wide Association Studies (GWAS) breast cancer susceptibility loci in Arabs: susceptibility and prognostic implications in Tunisians. *Breast Cancer Res Treat,* 135**,** 715-24.

Shore, P. (2005) A role for Runx2 in normal mammary gland and breast cancer bone metastasis. *J Cell Biochem,* 96**,** 484-9.

Slattery, M. L., Baumgartner, K. B., Giuliano, A. R., Byers, T., Herrick, J. S. & Wolff, R. K. (2011) Replication of five GWAS-identified loci and breast cancer risk among Hispanic and non-Hispanic white women living in the Southwestern United States. *Breast Cancer Res Treat,* 129**,** 531-9.

Smith, J., Bibikova, M., Whitby, F. G., Reddy, A. R., Chandrasegaran, S. & Carroll, D. (2000) Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains. *Nucleic Acids Res,* 28**,** 3361-9.

Soldner, F., Laganiere, J., Cheng, A. W., Hockemeyer, D., Gao, Q., Alagappan, R., Khurana, V., Golbe, L. I., Myers, R. H., Lindquist, S., Zhang, L., Guschin, D., Fong, L. K., Vu, B. J., Meng, X., Urnov, F. D., Rebar, E. J., Gregory, P. D., Zhang, H. S. & Jaenisch, R. (2011) Generation of isogenic pluripotent stem cells differing exclusively at two early onset Parkinson point mutations. *Cell,* 146**,** 318-31.

Sorlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Thorsen, T., Quist, H., Matese, J. C., Brown, P. O., Botstein, D., Lonning, P. E. & Borresen-Dale, A. L. (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A,* 98**,** 10869-74.

Sorrell, D. A. & Kolb, A. F. (2005) Targeted modification of mammalian genomes. *Biotechnol Adv,* 23**,** 431-69.

Soule, H. D., Maloney, T. M., Wolman, S. R., Peterson, W. D., Jr., Brenz, R., McGrath, C. M., Russo, J., Pauley, R. J., Jones, R. F. & Brooks, S. C. (1990) Isolation and characterization of a spontaneously immortalized human breast epithelial cell line, MCF-10. *Cancer Res,* 50**,** 6075-86.

Soule, H. D., Vazguez, J., Long, A., Albert, S. & Brennan, M. (1973) A human cell line from a pleural effusion derived from a breast carcinoma. *J Natl Cancer Inst,* 51**,** 1409-16.

Spinola, M., Leoni, V. P., Tanuma, J., Pettinicchio, A., Frattini, M., Signoroni, S., Agresti, R., Giovanazzi, R., Pilotti, S., Bertario, L., Ravagnani, F. & Dragani, T. A. (2005) FGFR4 Gly388Arg polymorphism and prognosis of breast and colorectal cancer. *Oncol Rep,* 14**,** 415-9.

Spivakov, M., Akhtar, J., Kheradpour, P., Beal, K., Girardot, C., Koscielny, G., Herrero, J., Kellis, M., Furlong, E. E. & Birney, E. (2012) Analysis of variation at transcription factor binding sites in Drosophila and humans. *Genome Biol,* 13**,** R49.

Steinberg, F., Zhuang, L., Beyeler, M., Kalin, R. E., Mullis, P. E., Brandli, A. W. & Trueb, B. (2009) The FGFRL1 receptor is shed from cell membranes, binds fibroblast growth factors (FGFs), and antagonizes FGF signaling in Xenopus embryos. *J Biol Chem,* 285**,** 2193-202.

Stewart, J., King, R., Hayward, J. & Rubens, R. (1982) Estrogen and progesterone receptors: correlation of response rates, site and timing of receptor analysis. *Breast Cancer Res Treat,* 2**,** 243-50.

Stinson, S., Lackner, M. R., Adai, A. T., Yu, N., Kim, H. J., O'Brien, C., Spoerke, J., Jhunjhunwala, S., Boyd, Z., Januario, T., Newman, R. J., Yue, P., Bourgon, R., Modrusan, Z., Stern, H. M., Warming, S., de Sauvage, F. J., Amler, L., Yeh, R. F. & Dornan, D. (2011) TRPS1 targeting by miR-221/222 promotes the epithelial-to-mesenchymal transition in breast cancer. *Sci Signal,* 4**,** ra41.

Sun, S., Jiang, Y., Zhang, G., Song, H., Zhang, X., Zhang, Y., Liang, X., Sun, Q. & Pang, D. (2012) Increased expression of fibroblastic growth factor receptor 2 is correlated with poor prognosis in patients with breast cancer. *J Surg Oncol,* 105**,** 773-9.

Sur, S., Pagliarini, R., Bunz, F., Rago, C., Diaz, L. A., Jr., Kinzler, K. W., Vogelstein, B. & Papadopoulos, N. (2009) A panel of isogenic human cancer cells suggests a therapeutic approach for cancers with inactivated p53. *Proc Natl Acad Sci U S A,* 106**,** 3964-9.

Taillon-Miller, P., Gu, Z., Li, Q., Hillier, L. & Kwok, P. Y. (1998) Overlapping genomic sequences: a treasure trove of single-nucleotide polymorphisms. *Genome Res,* 8**,** 748-54.

Takahashi, Y., Rayman, J. B. & Dynlacht, B. D. (2000) Analysis of promoter binding by the E2F and pRB families in vivo: distinct E2F proteins mediate activation and repression. *Genes Dev,* 14**,** 804-16.

Tannheimer, S. L., Rehemtulla, A. & Ethier, S. P. (2000) Characterization of fibroblast growth factor receptor 2 overexpression in the human breast cancer cell line SUM-52PE. *Breast Cancer Res,* 2**,** 311-20.

Tham, Y. L., Sexton, K., Kramer, R., Hilsenbeck, S. & Elledge, R. (2006) Primary breast cancer phenotypes associated with propensity for central nervous system metastases. *Cancer,* 107**,** 696-704.

Theodorou, V., Kimm, M. A., Boer, M., Wessels, L., Theelen, W., Jonkers, J. & Hilkens, J. (2007) MMTV insertional mutagenesis identifies genes, gene families and pathways involved in mammary cancer. *Nat Genet,* 39**,** 759-69.

Thompson, D. & Easton, D. (2004) The genetic epidemiology of breast cancer genes. *J Mammary Gland Biol Neoplasia,* 9**,** 221-36.

Thurman, R. E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M. T., Haugen, E., Sheffield, N. C., Stergachis, A. B., Wang, H., Vernot, B., Garg, K., John, S., Sandstrom, R., Bates, D., Boatman, L., Canfield, T. K., Diegel, M., Dunn, D., Ebersol, A. K., Frum, T., Giste, E., Johnson, A. K., Johnson, E. M., Kutyavin, T., Lajoie, B., Lee, B. K., Lee, K., London, D., Lotakis, D., Neph, S., Neri, F., Nguyen, E. D., Qu, H., Reynolds, A. P., Roach, V., Safi, A., Sanchez, M. E., Sanyal, A., Shafer, A., Simon, J. M., Song, L., Vong, S., Weaver, M., Yan, Y., Zhang, Z., Zhang, Z., Lenhard, B., Tewari, M., Dorschner, M. O., Hansen, R. S., Navas, P. A., Stamatoyannopoulos, G., Iyer, V. R., Lieb, J. D., Sunyaev, S. R., Akey, J. M., Sabo, P. J., Kaul, R., Furey, T. S., Dekker, J., Crawford, G. E. & Stamatoyannopoulos, J. A. (2012) The accessible chromatin landscape of the human genome. *Nature,* 489**,** 75-82.

Tirosh, I., Weinberger, A., Bezalel, D., Kaganovich, M. & Barkai, N. (2008) On the relation between promoter divergence and gene expression evolution. *Mol Syst Biol,* 4**,** 159.

Turner, N. & Grose, R. (2010) Fibroblast growth factor signalling: from development to cancer. *Nat Rev Cancer,* 10**,** 116-29.

Turner, N., Lambros, M. B., Horlings, H. M., Pearson, A., Sharpe, R., Natrajan, R., Geyer, F. C., van Kouwenhove, M., Kreike, B., Mackay, A., Ashworth, A., van de Vijver, M. J. & Reis-Filho, J. S. (2010) Integrative molecular profiling of triple negative breast cancers identifies amplicon drivers and potential therapeutic targets. *Oncogene,* 29**,** 2013-23.

Udler, M. S., Meyer, K. B., Pooley, K. A., Karlins, E., Struewing, J. P., Zhang, J., Doody, D. R., MacArthur, S., Tyrer, J., Pharoah, P. D., Luben, R., Bernstein, L., Kolonel, L. N., Henderson, B. E., Le Marchand, L., Ursin, G., Press, M. F., Brennan, P., Sangrajrang, S., Gaborieau, V., Odefrey, F., Shen, C. Y., Wu, P. E., Wang, H. C., Kang, D., Yoo, K. Y., Noh, D. Y., Ahn, S. H., Ponder, B. A., Haiman, C. A., Malone, K. E., Dunning, A. M., Ostrander, E. A. & Easton, D. F. (2009) FGFR2 variants and breast cancer risk: fine-scale mapping using African American studies and analysis of chromatin conformation. *Hum Mol Genet,* 18**,** 1692-703.

Urnov, F. D., Miller, J. C., Lee, Y. L., Beausejour, C. M., Rock, J. M., Augustus, S., Jamieson, A. C., Porteus, M. H., Gregory, P. D. & Holmes, M. C. (2005) Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature,* 435**,** 646-51.

Varley, J. M. (2003) Germline TP53 mutations and Li-Fraumeni syndrome. *Hum Mutat,* 21**,** 313-20.

Veltmaat, J. M., Relaix, F., Le, L. T., Kratochwil, K., Sala, F. G., van Veelen, W., Rice, R., Spencer-Dene, B., Mailleux, A. A., Rice, D. P., Thiery, J. P. & Bellusci, S. (2006) Gli3-mediated somitic Fgf10 expression gradients are required for the induction and patterning of mammary epithelium along the embryonic axes. *Development,* 133**,** 2325-35.

Volpi, E. V., Chevret, E., Jones, T., Vatcheva, R., Williamson, J., Beck, S., Campbell, R. D., Goldsworthy, M., Powis, S. H., Ragoussis, J., Trowsdale, J. & Sheer, D. (2000) Large-scale chromatin organization of the major histocompatibility complex and other regions of human chromosome 6 and its response to interferon in interphase nuclei. *J Cell Sci,* 113 (Pt 9)**,** 1565-76.

Walsh, T., Casadei, S., Coats, K. H., Swisher, E., Stray, S. M., Higgins, J., Roach, K. C., Mandell, J., Lee, M. K., Ciernikova, S., Foretova, L., Soucek, P. & King, M. C. (2006) Spectrum of mutations in BRCA1, BRCA2, CHEK2, and TP53 in families at high risk of breast cancer. *JAMA,* 295**,** 1379-88.

Wang, Y., Cortez, D., Yazdi, P., Neff, N., Elledge, S. J. & Qin, J. (2000) BASC, a super complex of BRCA1-associated proteins involved in the recognition and repair of aberrant DNA structures. *Genes Dev,* 14**,** 927-39.

Wang, Y. Z., Han, Y. S., Ma, Y. S., Jiang, J. J., Chen, Z. X., Wang, Y. C., Che, W., Zhang, F., Xia, Q. & Wang, X. F. (2012) Differential gene expression of Wnt signaling pathway in benign, premalignant, and malignant human breast epithelial cells. *Tumour Biol*.

Ward, T. M., Iorns, E., Liu, X., Hoe, N., Kim, P., Singh, S., Dean, S., Jegg, A. M., Gallas, M., Rodriguez, C., Lippman, M., Landgraf, R. & Pegram, M. D. (2012) Truncated p110 ERBB2 induces mammary epithelial cell migration, invasion and orthotopic

xenograft formation, and is associated with loss of phosphorylated STAT5. *Oncogene*.

Wenger, S. L., Senft, J. R., Sargent, L. M., Bamezai, R., Bairwa, N. & Grant, S. G. (2004) Comparison of established cell lines at different passages by karyotype and comparative genomic hybridization. *Biosci Rep,* 24**,** 631-9.

White, E. (1987) Projected changes in breast cancer incidence due to the trend toward delayed childbearing. *Am J Public Health*.

WHO, W. H. O. (2008) Globocan 2008. *http://globocan.iarc.fr/factsheets/populations/factsheet.asp?uno=900#WOMEN,* 21.10.10.

Wong, E. S., Fong, C. W., Lim, J., Yusoff, P., Low, B. C., Langdon, W. Y. & Guy, G. R. (2002) Sprouty2 attenuates epidermal growth factor receptor ubiquitylation and endocytosis, and consequently enhances Ras/ERK signalling. *EMBO J,* 21**,** 4796-808.

Wooster, R., Bignell, G., Lancaster, J., Swift, S., Seal, S., Mangion, J., Collins, N., Gregory, S., Gumbs, C. & Micklem, G. (1995) Identification of the breast cancer susceptibility gene BRCA2. *Nature,* 378**,** 789-92.

Wray, G. A., Hahn, M. W., Abouheif, E., Balhoff, J. P., Pizer, M., Rockman, M. V. & Romano, L. A. (2003) The evolution of transcriptional regulation in eukaryotes. *Mol Biol Evol,* 20**,** 1377-419.

Xu, F., You, X., Liu, F., Shen, X., Yao, Y., Ye, L. & Zhang, X. (2013) The oncoprotein HBXIP up-regulates Skp2 via activating transcription factor E2F1 to promote proliferation of breast cancer cells. *Cancer Lett*.

Yan, H., Yuan, W., Velculescu, V. E., Vogelstein, B. & Kinzler, K. W. (2002) Allelic variation in human gene expression. *Science,* 297**,** 1143.

Yarden, Y. (2001) Biology of HER2 and its importance in breast cancer. *Oncology,* 61 Suppl 2**,** 1-13.

Yu, J., Wang, F., Yang, G. H., Wang, F. L., Ma, Y. N., Du, Z. W. & Zhang, J. W. (2006) Human microRNA clusters: genomic organization and expression profile in leukemia cell lines. *Biochem Biophys Res Commun,* 349**,** 59-68.

Zhang, K., Li, J. B., Gao, Y., Egli, D., Xie, B., Deng, J., Li, Z., Lee, J. H., Aach, J., Leproust, E. M., Eggan, K. & Church, G. M. (2009) Digital RNA allelotyping reveals tissue-specific and allele-specific gene expression in human. *Nat Methods,* 6**,** 613-8.

Zhang, S. M., Lee, I. M., Manson, J. E., Cook, N. R., Willett, W. C. & Buring, J. E. (2007) Alcohol consumption and breast cancer risk in the Women's Health Study. *Am J Epidemiol,* 165**,** 667-76.

Zhang, X., Ibrahimi, O. A., Olsen, S. K., Umemori, H., Mohammadi, M. & Ornitz, D. M. (2006) Receptor specificity of the fibroblast growth factor family. The complete mammalian FGF family. *J Biol Chem,* 281**,** 15694-700.

Zhao, J. J., Lin, J., Yang, H., Kong, W., He, L., Ma, X., Coppola, D. & Cheng, J. Q. (2008) MicroRNA-221/222 negatively regulates estrogen receptor alpha and is associated with tamoxifen resistance in breast cancer. *J Biol Chem,* 283**,** 31079-86.

Zhao, Z., Yu, N., Fu, Y. X. & Li, W. H. (2006) Nucleotide variation and haplotype diversity in a 10-kb noncoding region in three continental human populations. *Genetics,* 174**,** 399-409.

Zheng, W., Cai, Q., Signorello, L. B., Long, J., Hargreaves, M. K., Deming, S. L., Li, G., Li, C., Cui, Y. & Blot, W. J. (2009) Evaluation of 11 breast cancer susceptibility loci in African-American women. *Cancer Epidemiol Biomarkers Prev,* 18**,** 2761-4.

Zhu, X., Asa, S. L. & Ezzat, S. (2009) Histone-acetylated control of fibroblast growth factor receptor 2 intron 2 polymorphisms and isoform splicing in breast cancer. *Mol Endocrinol,* 23**,** 1397-405.